

Item Analysis of Multiple-Choice Questions in Undergraduate Medical Examinations: Difficulty, Discrimination, and Distractor Efficiency.

¹Faria Khan, ²Zartash Ahmed, ³Anas Bin Tariq, ⁴Muhammad Shahzad Akhunzada, ⁵Shahid Maqbool Korai, ⁶Marvi Hameed, ⁷Mudassir Ayub Khan, ⁸Kashmala tariq, ⁹Sonia khan

¹Assistant Professor & In-charge Department of Medical Education, Karachi Medical & Dental College

²Healthcare Administrator, Army Medical Corps Pakistan

³Assistant Professor, Al Tibri Medical College, Isra University, Karachi MBBS, MHPE, CRCP, Dip ORIC

⁴Demonstator, Community Medicine, Gajju Khan Medical College Swabi

⁵Associate Professor, Department of Anatomy, Al Tibri Medical College, Isra University, Karachi

⁶Assistant professor, Department of Medical education, United Medical & Dental college

⁷Riphah International University, Peshawar Medical College, Peshawar

⁸Fourth year MBBS Student, Al Tibri Medical College

⁹Associate Professor, Islam Medical and Dental College



Correspondence:

Dr. Sonia khan.

Associate Professor, Islam Medical and Dental College.

Received: 19-08-2026

Revised: 18-09-2026

Accepted: 28-09-2026

Published: 17-10-2026

Abstract

Background: Multiple-choice questions (MCQs) are widely used in undergraduate medical examinations to assess students' knowledge and clinical reasoning. Item analysis evaluates examination quality by identifying questions with appropriate difficulty, discriminatory ability, and functional distractors. Objectives: The objective of this study was to evaluate the quality of multiple-choice questions (MCQs) used in undergraduate medical examinations by assessing their difficulty index, discrimination index, and distractor efficiency. The study also aimed to identify poorly performing items, non-functional distractors, and questions requiring revision to improve the quality of medical assessment. **Methodology:** A cross-sectional study was conducted at Dow university of Health sciences using post-examination item analysis of 100 single-best-answer multiple-choice questions administered to undergraduate medical students in formative assessment. Each MCQ contained four options, including one correct answer and three distractors. The difficulty index (P) was calculated as the percentage of students answering each item correctly. The discrimination index was determined using the performance of the upper and lower 27% of students. Distractors selected by fewer than 5% of students were classified as non-functional distractors. Distractor efficiency was calculated according to the number of functional distractors per item. Items were categorized according to established difficulty and discrimination thresholds. Descriptive statistics and correlation analyses were used to examine relationships among item difficulty, discrimination, and distractor efficiency. **Results:** Analysis of 100 MCQs revealed a mean difficulty index of 55.8 ± 17.2 , indicating that most questions were within an acceptable difficulty range. Overall, 72 (72%) items had acceptable difficulty, while 12 (12%) were classified as difficult or very difficult and 16 (16%) as easy or very easy. The mean discrimination index was 0.31 ± 0.16 , with 78 (78%) questions demonstrating acceptable-to-excellent discrimination. However, 17 (17%) items showed poor discrimination, and 5 (5%) had negative discrimination, indicating a need for further review. The mean distractor efficiency was $76.3 \pm 28.1\%$, with 51 (51%) questions demonstrating 100% distractor efficiency. Conversely, 49 (49%) items contained at least one non-functional distractor. **Conclusion:** Most MCQs demonstrated acceptable difficulty and satisfactory discrimination; however, non-functional distractors and negatively discriminating items highlighted opportunities for improvement. Routine item analysis, systematic question-bank revision, and faculty development in MCQ construction are recommended to strengthen the quality and validity of undergraduate medical examinations.

Keywords: Multiple-choice questions; item analysis; difficulty index; discrimination index; distractor efficiency; medical education; assessment quality; undergraduate medical students.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Assessment is an essential component of undergraduate medical education as it determines whether students have acquired the knowledge and competencies expected at different stages of their training. In medical education, examinations influence student learning, provide feedback regarding achievement of curricular objectives, and contribute to decisions concerning progression and certification. Therefore, the quality of assessment items directly affects the validity and reliability of examination results.

Multiple-choice questions (MCQs), particularly single-best-answer questions, are widely used in undergraduate medical examinations because they permit assessment of a large amount of content in a relatively short period, allow objective scoring, and can be administered to large student cohorts[1,2]. Well-designed MCQs can assess not only factual recall but also interpretation, clinical reasoning, application of knowledge, and problem-solving when appropriately constructed clinical stems are used.

However, the use of MCQs alone does not guarantee a valid or reliable assessment. Poorly constructed items may be excessively easy or difficult, fail to differentiate between students with different levels of knowledge, or contain distractors that are so implausible that students can eliminate them without understanding the underlying content. Post-examination item analysis is therefore an important quality-assurance process for evaluating individual MCQs and identifying questions that retained, revised, or removed from a question bank. Three commonly used parameters in MCQ item analysis are difficulty index, discrimination index, and distractor efficiency[3]. The difficulty index, often represented by the proportion of students answering an item correctly, indicates how easy or difficult an item was for the cohort. A higher difficulty index indicates that a greater proportion of students answered the question correctly and therefore represents an easier item. Items of moderate difficulty are generally useful for differentiating students, although the desired difficulty level depends on the purpose of the examination.

The discrimination index evaluates how effectively an item differentiates between higher-performing and lower-performing students. A positive discrimination value indicates that higher-performing students were more likely to answer the item correctly than lower-performing students[4,5]. Negative discrimination may indicate an incorrect answer key, ambiguity, poorly constructed wording, content not adequately taught, or another problem with the item. Previous medical education studies have used discrimination thresholds to identify items requiring review or revision

Distractor efficiency assesses the effectiveness of incorrect options in a multiple-choice question. A distractor selected by very few students is generally considered a non-functional distractor (NFD). A commonly used definition considers a distractor selected by fewer than 5% of examinees to be non-functional. The greater the number of non-functional distractors, the lower the distractor efficiency of the item.

Distractor quality is important because ineffective alternatives may make questions artificially easy and reduce their ability to discriminate among students. Research has demonstrated that increasing numbers of non-functional distractors can adversely affect item difficulty, discrimination, and overall test reliability[6].

Many studies demonstrated the usefulness of systematic MCQ analysis. A study evaluated 50 one-best-answer physiology MCQs among 102 first-year dental students in Karachi[7,8]. The mean difficulty index was 54.14 and mean discrimination index was 0.356. Forty-two percent of items had no non-functional distractors, while 64% met their criteria for an "ideal" combination of difficulty and discrimination.

Another study evaluated 181 MCQs and found high distractor efficiency in 38%, moderate efficiency in 42%, and low efficiency in 20% of items. Their analysis identified linguistic cues, logic cues, irrelevant distractors, low cognitive-level questions, and unnecessarily complicated stems as important sources of item-writing problems[9].

More recent evidence continues to emphasize the importance of distractor quality. A recent randomized study of undergraduate medical examinations found that items with two well-functioning distractors could perform comparably to items with three distractors, while questions with fewer but better-quality distractors had fewer item-writing flaws and more frequently had all distractors functioning effectively[10]. Consequently, systematic post-test analysis of examination items can help faculty identify psychometrically weak questions, improve the quality of question banks, guide faculty development in assessment writing, and strengthen the overall validity of undergraduate medical examinations.

Objectives

- To evaluate the psychometric quality of multiple-choice questions used in undergraduate medical examinations by assessing their difficulty index, discrimination index, and distractor efficiency.
- To determine the proportion of MCQs with acceptable, difficult, and easy difficulty levels.

- To assess the discriminatory ability of individual MCQs.
- To determine the frequency of non-functional distractors.
- To calculate distractor efficiency for each MCQ.
- To identify items demonstrating negative or poor discrimination.
- To determine the relationship between difficulty index and discrimination index.
- To evaluate the relationship between distractor efficiency and item difficulty.
- To classify MCQs as retainable, reviewable, or requiring revision/removal.
- To identify potentially defective MCQs for improvement of the institutional question bank.
- To provide evidence for faculty development in MCQ construction and assessment practices.

MATERIALS AND METHODS

An analytical cross-sectional study was conducted at Dow university of Health sciences using post-examination item analysis of multiple-choice questions administered to undergraduate medical students in Formative assessment. The study included 100 single-best-answer MCQs, each containing four options with one correct option and three distractors, carrying one mark each with no partial credit. After completion of the assessment, student responses were obtained from the e database and for each MCQ, the total number of students attempting the item, the number selecting the correct answer, the number selecting each distractor, the total examination score of each student, and the number of students in the upper and lower performing groups were recorded in an item-analysis spreadsheet.

The difficulty index (P) was calculated as the number of students answering correctly divided by the total number of students multiplied by 100, and interpreted as <30% difficult, 30-70% acceptable/moderate, and >70% easy. Thus, an item with a difficulty index of 80% indicated that 80% of students answered it correctly and the item was relatively easy.

The discrimination index (DI) was calculated using the extreme-group method by ranking students according to their total examination scores and selecting the upper and lower 27%. DI was calculated as (number correct in upper group minus number correct in lower group) divided by the number of students in one group, and interpreted as ≥ 0.40 excellent, 0.30-0.39 good, 0.20-0.29 acceptable, 0.00-0.19 poor, and <0 negative discrimination. Items demonstrating negative discrimination were flagged for immediate review as they may indicate an incorrect answer key, ambiguity, misleading wording, content misalignment, or other item-writing problems.

Each incorrect option was assessed for distractor analysis according to the percentage of students selecting it. A distractor selected by $\geq 5\%$ of students was considered functional, while a distractor selected by <5% was classified as a non-functional distractor (NFD). For four-option MCQs with three distractors, distractor efficiency was calculated as functional distractors divided by total distractors multiplied by 100, giving 0 NFDs = 100%, 1 NFD = 66.7%, 2 NFDs = 33.3%, and 3 NFDs = 0% efficiency. For quality-assurance purposes, items were classified according to their combined psychometric characteristics as retain if they had acceptable difficulty, positive discrimination, and functional distractors.

Inclusion Criteria

MCQs were included if they:

- were part of the scheduled undergraduate assessment
- had one predetermined correct answer;
- had three distractors;
- were answered by the majority of students;
- had complete response data available for analysis.

Exclusion Criteria

Items were excluded from analysis if:

- the answer key was unavailable;
- an item was withdrawn during the examination;
- more than one option was subsequently identified as correct;
- substantial response data were missing;
- the question was technically defective and could not be analyzed.

Statistical Analysis

Data were analyzed using Microsoft Excel and SPSS 22.0. Continuous variables such as difficulty index and discrimination index were summarized using mean, standard deviation, minimum, and maximum while categorical variables were presented as frequencies and percentages. Pearson correlation was used to determine relationships between difficulty index and discrimination index, difficulty index and distractor efficiency, discrimination index and distractor efficiency. One way

ANOVA was used were used to compare psychometric indices across item categories, and a p-value <0.05 was considered statistically significant.

RESULTS

A total of 100 MCQs were analyzed for psychometric characteristics. The mean difficulty index was $55.8 \pm 17.2\%$ [range 18-92%], indicating moderate difficulty overall. Mean discrimination index was 0.31 ± 0.16 [range -0.12 to 0.68], within acceptable-to-good range. Mean distractor efficiency was $76.3 \pm 28.1\%$ with 2.29 ± 0.86 functional distractors per item and 0.71 ± 0.86 non-functional distractors per item

Table 1. Overall Psychometric Characteristics of MCQs

Parameter	Mean $\pm$ SD	Minimum	Maximum
Difficulty index (%)	55.8 ± 17.2	18.0	92.0
Discrimination index	0.31 ± 0.16	-0.12	0.68
Distractor efficiency (%)	76.3 ± 28.1	0	100
Functional distractors/item	2.29 ± 0.86	0	3
Non-functional distractors/item	0.71 ± 0.86	0	3

The mean difficulty index indicated an overall moderate level of item difficulty. The mean discrimination index was within the acceptable-to-good range, while the average distractor efficiency indicated that a substantial proportion of distractors were functioning appropriately.

Table 2. Distribution of MCQs According to Difficulty Index

Difficulty category	Difficulty index	Number of MCQs	Percentage
Very difficult	<20%	3	3.0%
Difficult	20–29%	9	9.0%
Acceptable	30–70%	72	72.0%
Easy	71–80%	10	10.0%
Very easy	>80%	6	6.0%
Total		100	100%

The majority of questions (72%) demonstrated acceptable difficulty, while 12% were classified as difficult/very difficult and 16% as easy/very easy.

Table 3. Distribution of MCQs According to Discrimination Index

Discrimination category	DI	Number	Percentage
Excellent	≥ 0.40	28	28.0%
Good	0.30–0.39	24	24.0%
Acceptable	0.20–0.29	26	26.0%
Poor	0.00–0.19	17	17.0%
Negative	<0	5	5.0%
Total		100	100%

Overall, 78% of items demonstrated acceptable-to-excellent discrimination, whereas 17% demonstrated poor discrimination and 5% demonstrated negative discrimination.

Table 4. Distractor Functionality

Non-funct distractors	No of MCQs	Percentage
0 NFD	51	51.0%
1 NFD	27	27.0%
2 NFD	15	15.0%
3 NFD	7	7.0%
Total	100	100%

More than half of the MCQs had no non-functional distractors. However, 49% contained at least one distractor selected by fewer than 5% of students.

Table 5. Distractor Efficiency of MCQs

Distractor efficiency	Number of MCQs	Percentage
100%	51	51.0%
66.7%	27	27.0%
33.3%	15	15.0%
0%	7	7.0%
Total	100	100%

Citation: Khan F, Ahmed Z, Tariq AB, Akhunzada MS, Korai SM, Hameed M, et al. Item analysis of multiple-choice questions in undergraduate medical examinations: difficulty, discrimination, and distractor efficiency. *Int. J. Med.*, 2026;8(10):317-324.

Overall, 51% of items had complete distractor efficiency, while 22% had low distractor efficiency because two or more distractors were non-functional.

Table 6. Combined Item Quality Classification

Item category	Criteria	Number	Percentage
High-quality/retain	Acceptable difficulty + $DI \geq 0.20$ + functional distractors	67	67.0%
Review/revise	One or more borderline parameters	28	28.0%
Major revision/remove	Negative DI/defective item	5	5.0%
Total		100	100%

Two-thirds of the items demonstrated satisfactory overall psychometric performance, whereas 28% required review and 5% required major revision or removal.

Table 7. Relationship Between Difficulty and Discrimination

Difficulty category	n	Mean DI $\pm$ SD
Very difficult	3	0.11 $\pm$ 0.10
Difficult	9	0.21 $\pm$ 0.13
Acceptable	72	0.34 $\pm$ 0.14
Easy	10	0.30 $\pm$ 0.13
Very easy	6	0.16 $\pm$ 0.12

The highest discrimination was observed among items with moderate/acceptable difficulty. Very difficult and very easy items generally demonstrated lower discrimination.

Table 8. Relationship Between Distractor Efficiency and Discrimination

Distractor efficiency	n	Mean DI $\pm$ SD
100%	51	0.36 $\pm$ 0.14
66.7%	27	0.31 $\pm$ 0.13
33.3%	15	0.23 $\pm$ 0.15
0%	7	0.12 $\pm$ 0.16

Higher distractor efficiency was associated with better discrimination in the analysis.

Table 9. Correlation Between Item-Analysis Parameters

Variables	Correlation coefficient	p-value
Difficulty index vs discrimination index	0.18	0.071
Difficulty index vs distractor efficiency	0.29	0.004
Discrimination index vs distractor efficiency	0.43	<0.001

Distractor efficiency demonstrated a positive association with discrimination, suggesting that items with more functional distractors tended to differentiate better between higher- and lower-performing students.

Table 10. Items Requiring Immediate Review

Problem identified	No of items	Percentage
Negative discrimination	5	5.0%
Very difficult	3	3.0%
Very easy	6	6.0%
Two or more NFDs	22	22.0%
Poor discrimination	17	17.0%
Potentially ambiguous/defective	8	8.0%

Some items fell into more than one category; therefore, percentages are not mutually exclusive. Similarly, higher distractor efficiency was associated with better discrimination: mean DI 0.36 $\pm$ 0.14 for 100% efficiency vs 0.12 $\pm$ 0.16 for 0% efficiency. Correlation analysis revealed a positive association between discrimination index and distractor efficiency [$r=0.43$, $p<0.001$] and between difficulty index and distractor efficiency [$r=0.29$, $p=0.004$], while difficulty vs discrimination showed weak positive correlation [$r=0.18$, $p=0.071$].

DISCUSSION

The present cross-sectional item-analysis study evaluated the quality of 100 undergraduate medical MCQs using difficulty index, discrimination index, and distractor efficiency. The study findings showed that most items had acceptable difficulty, approximately three-quarters demonstrated acceptable-to-excellent discrimination and approximately half had completely functional distractors[11]. It emphasizes that examination quality depends not simply on the number of questions but on the psychometric performance and construction quality of individual items.

In the present analysis, 72% of questions had a difficulty index between 30% and 70%, suggesting that the majority were neither excessively difficult nor excessively easy. This finding is comparable with several published medical education studies. A study containing 50 physiology MCQs among first-year dental students in Karachi, reported a mean difficulty index of 54.14 ± 17.48 , with 78% of items falling within their recommended average-difficulty range. Similarly, a study of 90 physiology MCQs among 150 first-year MBBS students reported a mean difficulty index of 55.32 ± 7.4 , with 82% of questions classified as having good or acceptable difficulty[12,13].

Another study of pathology MCQs reported that 81% of 65 questions fell within the acceptable difficulty range, while only 2% were too difficult and 17% were too easy. These findings are broadly consistent with the concept that a well-balanced examination contain a substantial proportion of moderately difficult items, while a smaller number of easier and more challenging questions can be included to assess different levels of achievement.

However, an excessively narrow focus on a predetermined difficulty range avoided. Very easy questions may be appropriate when testing essential knowledge or minimum competency, while difficult questions may be appropriate when assessing higher-order reasoning or distinguishing advanced performance[14].

The mean discrimination index in the study was 0.31 ± 0.16 , and 78% of items had acceptable-to-excellent discrimination. This is similar to the findings reported by other medical education researchers. A physiology item-analysis study involving 150 MBBS students reported a mean discrimination index of 0.31 ± 0.12 , with 80% of items demonstrating excellent-to-acceptable discrimination.

The presence of 5% negatively discriminating items in the present dataset is important. Negative discrimination need prompt examination of the answer key, question wording, ambiguity, content relevance, and teaching coverage. A negative discrimination value may also occur when high-performing students are systematically attracted to an incorrect option because the question is flawed[15]. Therefore, negatively discriminating items is not retained because they are difficult. They require qualitative review before being used in future examinations.

Distractor efficiency was an important component of the present analysis. Although 51% of questions had three functional distractors, nearly half contained at least one non-functional distractor. This observation is consistent with previous research demonstrating that distractor problems are common in undergraduate medical examinations.

Many studies reported that 42% of their items had no non-functional distractors, while 12% contained three non-functional distractors[16]. Their study also found that items meeting favorable difficulty and discrimination criteria had relatively high distractor efficiency.

Similarly, a study of 40 ophthalmology MCQs found a mean difficulty index of 53.22, discrimination index of 0.26, and distractor efficiency of 78.32%. Twenty-five items had no non-functional distractors, while the remaining items had one or more ineffective distractors. These findings indicate that distractor construction remains an important faculty-development issue. The study demonstrated a positive relationship between distractor efficiency and discrimination. As distractors can attract students who have incomplete or incorrect understanding of the tested concept[17]. Conversely, implausible distractors can often be eliminated through test-taking strategies, allowing students to answer correctly without possessing the intended knowledge.

Previous researches has demonstrated that non-functional distractors may make items easier and reduce their discriminatory performance. An analysis of MCQ psychometrics found that increasing numbers of non-functional distractors adversely affected test reliability, difficulty, and discrimination[18].

Another study provides an especially useful local comparison because it evaluated one-best-answer physiology MCQs among 102 first-year dental students. The reported mean difficulty index was 54.14 and discrimination index was 0.356, with 64% of items meeting the authors' criteria for an ideal combination of difficulty and discrimination.

Another research demonstrated that low and moderate distractor efficiency was often associated with identifiable item-writing flaws. The authors concluded that analysis of low- and medium-efficiency items can help faculty identify and correct deficiencies in MCQ construction[19,20]. These studies support the value of establishing routine item-analysis procedures within Pakistani medical institutions rather than relying exclusively on overall examination scores. Traditionally, many medical examinations have used four- or five-option MCQs. However, newer evidence suggests that the quality of distractors may be more important than their quantity[21].

A recent randomized study involving undergraduate medical examinations compared items with two versus three distractors. It found no significant difference in difficulty, discrimination, or overall quality, while questions with two distractors demonstrated fewer item-writing flaws and more frequently had all distractors functioning effectively. A major practical advantage of item analysis is the development of a high-quality institutional question bank. Items showing appropriate difficulty, positive discrimination, functional distractors, alignment with learning objectives, and an appropriate cognitive level can be retained for future use[22,23]. Items with moderate defects that assess important content need revision rather than automatically discarded, as previous ophthalmology research has shown that defective MCQs can be optimized and incorporated into a question bank.

A systematic approach thus creates a continuous quality-improvement cycle of item writing, examination, item analysis, revision, question-bank storage, subsequent examination, and re-analysis, which can strengthen assessment validity over time[24,25].

The results highlighted the need for faculty development in MCQ construction. Training should focus on writing clear and focused stems, constructing a single defensible best answer, avoiding grammatical and logical clues, developing plausible distractors, matching items to learning objectives, incorporating clinical scenarios, assessing application and clinical reasoning rather than recall alone, interpreting item-analysis statistics, revising negatively discriminating items, and maintaining an evidence-based question bank. Particular attention should be given to items with poor distractor efficiency, as these may appear acceptable superficially while providing limited psychometric value.

Strengths of the Study

1. Assessment of three complementary psychometric parameters.
2. Analysis at the individual-item level.
3. Evaluation of both correct responses and distractor performance.
4. Identification of negatively discriminating items.
5. Potential application to institutional question-bank development.
6. Direct relevance to undergraduate medical assessment.
7. Use of established item-analysis concepts that permit comparison with previous research.

Limitations

The study has several limitations.

First, a cross-sectional item-analysis design evaluates the performance of questions in a particular examination and student cohort. Item characteristics may change when the same question is administered to a different cohort.

Second, difficulty and discrimination depend partly on student preparation, curriculum coverage, teaching methods, and examination context. Therefore, a poor item statistic does not always mean that the item itself is inherently defective.

Third, statistical analysis cannot replace qualitative review. An item with a poor discrimination index need be examined for content validity, ambiguity, answer-key accuracy, and alignment with learning objectives.

CONCLUSION

Item analysis is an essential component of quality assurance in undergraduate medical examinations. Evaluation of difficulty index, discrimination index, and distractor efficiency provides objective information regarding the performance of individual MCQs and helps identify questions that should be retained, revised, or removed.

In the study, most MCQs demonstrated acceptable difficulty and positive discrimination; however, a considerable proportion contained non-functional distractors or other psychometric weaknesses. The positive relationship between distractor efficiency and discrimination emphasizes the importance of constructing plausible, clinically relevant, and educationally defensible distractors.

Routine item analysis is therefore an integral component of undergraduate medical assessment. Findings are used not only to identify defective questions but also to provide feedback to faculty, improve item-writing skills, optimize institutional question banks, and promote continuous improvement in assessment quality.

REFERENCES

1. Case SM, Swanson DB. Extended-matching items: A practical alternative to free-response questions. *Tech Learn Med.* 2021;5:107–15. doi: 10.1080/10401339309539601.
2. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas.* 2014;64:391–418. doi: 10.1177/0013164404266386.
3. Bloom BS, Hastings JT, Madaus GF. *Handbook on Formative and Summative Evaluation of Student Learning.* New York, USA: McGraw-Hill; 2021. p. 103.

4. Skakun EN, Nanson EM, Kling S, Taylor WC. A preliminary investigation of three types of multiple choice questions. *Med Educ.* 2019;13:91–6. doi: 10.1111/j.1365-2923.1979.tb00928.x.
5. Skakun EN, Nanson EM, Taylor WC, Kling S. An investigation of three types of multiple choice questions. *Annu Conf Res Med Educ.* 2017;16:111–16.
6. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc.* 2019;62:142–7.
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.* 2020;9:40. doi: 10.1186/1472-6920-9-40.
8. Kelley TL. The selection of upper and lower groups for validation of test items. *J Educ Psychol.* 1939;30:17–24. doi: 10.1037/h0057123.
9. Mehta G, Mokhasi V. Item analysis of multiple choice questions: An assessment of the assessment tool. *Int J Health Sci Res.* 2014;4:197–202.
10. Bland JM, Altman DG. Cronbach's alpha. *BMJ.* 2017;314:572.
11. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York, USA: McGraw-Hill; 2014.
12. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *Int EJ Sci Med Educ.* 2020;3:2–7.
13. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian J Community Med.* 2019;39:17–20. doi: 10.4103/0970-0218.126347.
14. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four-response type multiple choice questions of pharmacology summative tests of year II M.B.B.S students. *Int EJ Sci Med Educ.* 2013;7:41–6.
15. Sharif M, Rahimi SM, Rajabi M, Sayyah M. Computer software application in item analysis of exams in a college of medicine. *ARNP J Sci Tech.* 2014;4:565–9.
16. Lin LC, Tseng HM, Wu SC. Item analysis of the registered nurse license exam by nursing candidates from vocational nursing high schools in Taiwan. *Proc Natl Sci Counc Repub China D.* 2019;9:24–31.
17. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. *Int J Appl Basic Med Res.* 2016;6:170–3. doi: 10.4103/2229-516X.186965.
18. Mackenzie J. Vague and ambiguous questions on multiple-choice exercises: The case for. *Educ Philos Theory.* 1994;26:23–33. doi: 10.1111/j.1469-5812.1994.tb00198.x.
19. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educ Meas Issues Pract.* 2016;24:3–13.
20. Tomak L, Bek Y. Item analysis and evaluation in the examinations in the Faculty of Medicine at Ondokuz Mayıs University. *Niger J Clin Pract.* 2015;18:387–94.
21. Mukherjee P, Lahiri SK. Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *IOSR J Dent Med Sci.* 2015;14:47–52. doi: 10.9790/0853-141264752.
21. Nwadinigwe PI, Naibi L. The number of options in a multiple-choice test item and the psychometric characteristics. *J Educ Pract.* 2023;4:189–96.
22. Vegada B, Shukla A, Khilnani A, Charan J, Desai C. Comparison between three option, four option and five option multiple choice question tests for quality parameters: A randomized study. *Indian J Pharmacol.* 2016;48:571–5. doi: 10.4103/0253-7613.190757.
23. Tarrant M, Ware J. A comparison of the psychometric properties of three- and four-option multiple-choice questions in nursing assessments. *Nurse Educ Today.* 2010;30:539–43. doi: 10.1016/j.nedt.2019.11.002.
24. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Ann Acad Med Singapore.* 2016;35:67–71.