European Journal of Clinical Pharmacy
Print ISSN: 2385-409X **Online ISSN**: 3105-0409
Website: https://farmclin.com/

Original Research Article

Improving Assessment Quality in Medical Education: The Application of Item Analysis to Multiple-Choice Questions

Article History:

Name of Author:
Dr Sarah Arif[1], Dr Junaid Sarfaraz[2] and Dr Ayesha Junaid[3]

Affiliation: *[1]Professor, Department Of Pathology, Gomal Medical College, Dera Ismail Khan*
[2]Professor, Rector/Director Academics, Health Service Academy, Islamabad
[3]Assistant Professor of Linguistic, Department Of English Language & Literature, College of Sciences & Humanities, Prince Sattam Bin Abdul Aziz University

Corresponding Author:
Dr. Sarah Arif

Received: 20-01-2026
Revised: 02-02-2026
Accepted: 25-02-2026
Published: 03-03-2026

Abstract: ***Background***: Multiple-choice questions (MCQs) are widely used in medical education due to their objectivity and broad content coverage. However, poorly constructed items can reduce assessment validity and reliability. Item analysis using difficulty index, discrimination index, and distractor efficiency provides a systematic method to evaluate and improve the quality of examination items. Objective: To evaluate the quality of MCQs used in a formative medical examination through item analysis and to identify items that are good, acceptable, or require revision based on difficulty index, discrimination index, and distractor efficiency. ***Methodology*** A descriptive cross-sectional study was conducted to analyze MCQs from a formative undergraduate medical examination of 3rd Professional year MBBS conducted at Dow university of Health sciences anonymized response data were used to calculate difficulty index, discrimination index, and distractor efficiency. Items were categorized according to predefined criteria into good, acceptable, or needing revision. Data were analyzed using standard statistical formulas in spreadsheet software ***Results***: A study analyzing 100 multiple-choice questions from a formative assessment of 3rd-year MBBS students in Dow University of Health sciences revealed areas for improvement. While 62% of questions had optimal difficulty, 23% were too easy and 15% too hard. Discriminatory power was excellent in 28% of items, but 19% showed poor discrimination, indicating a need for revision. Distractor efficiency was suboptimal, with only 45% of questions having fully functional distractors. The assessment showed sound psychometric quality overall, but routine item analysis is crucial to identify and address weaknesses, ensuring high-quality evaluations. ***Conclusion***: Item analysis is an effective and practical tool for improving assessment quality in medical education. Regular post-examination analysis enhances validity, reliability, and fairness of MCQs, and supports continuous improvement in assessment practices within competency-based curricula.

Keywords: Multiple-choice questions (MCQs), Item analysis, Difficulty index, Discrimination index, Distractor efficiency, Assessment quality.

INTRODUCTION

Assessment is a cornerstone of medical education, serving not only to measure students' knowledge but also to evaluate higher-order cognitive skills such as clinical reasoning, problem-solving, and decision-making. In the context of competency-based medical education (CBME), assessment strategies must be valid, reliable, objective, and aligned with defined learning outcomes to ensure that graduates are adequately prepared for clinical practice[1].

Multiple-choice questions (MCQs) remain one of the most frequently employed assessment tools in undergraduate and postgraduate medical education[2]. Their widespread use in both formative and formative examinations is attributed to several advantages, including objectivity in scoring, high reliability, feasibility for large cohorts, and the ability to sample a wide content domain within a limited

time[3,4]. When appropriately constructed, MCQs can assess not only factual recall but also application, analysis, and clinical reasoning.

Despite these strengths, the quality of MCQs is highly dependent on sound item construction principles. Poorly written items—such as those with ambiguous stems, implausible distractors, or unintended cues can undermine assessment validity, compromise reliability, and introduce construct-irrelevant variance[5]. Such flaws may lead to inaccurate interpretation of students' performance, potentially disadvantaging learners and weakening the credibility of the assessment process.

Item analysis is a systematic post-examination statistical method used to evaluate the psychometric properties of individual test items. It provides objective evidence regarding item performance and overall test quality[6,7]. The difficulty index indicates the proportion of students who answered an item correctly, thereby identifying questions that are excessively easy or overly difficult. The discrimination index reflects an item's ability to differentiate between high-performing and low-performing students, serving as a key indicator of item effectiveness. Distractor efficiency assesses the functionality of incorrect options by determining whether they are selected by examinees, thus evaluating their plausibility and contribution to discrimination[8].

The routine application of item analysis plays a critical role in quality assurance of assessments. It enables educators to identify flawed items, refine or discard non-performing questions, and build a robust question bank of high-quality MCQs[9]. Furthermore, item analysis supports evidence-based decision-making in assessment design, promotes fairness and transparency, and enhances the overall reliability and validity of examinations.

The present study aims to evaluate the effectiveness of item analysis in improving the quality of MCQs used in a formative medical examination. By analyzing key item parameters i.e, difficulty index, discrimination index and distractor efficiency, this study seeks to provide empirical insights into the strengths and weaknesses of existing MCQs and highlight the value of item analysis as an essential tool in strengthening assessment practices in medical education.

OBJECTIVES:

The goal of the study was to evaluate the quality of multiple-choice questions (MCQs) used in a formative medical examination through item analysis. Specifically, it sought to calculate the difficulty index of MCQs, determine the discrimination index of each item, assess distractor efficiency, and categorize items as good, acceptable, or needing revision based on predefined criteria. The ultimate goal was to provide recommendations for improving assessment validity and reliability, thereby enhancing the overall quality of the examination. By doing so, the study aimed to contribute to the continuous improvement of assessment practices within competency-based medical curricula.

METHODOLOGY

This descriptive cross-sectional study was conducted to evaluate the quality of multiple-choice questions (MCQs) used in a formative medical examination of 3rd Professional year MBBS conducted at Dow university of Health sciences through item analysis. Undergraduate medical students who appeared in the formative assessment constituted the study population, and all completed answer sheets were included in the analysis. The examination comprised a single best answer MCQ paper with (number) items, each containing one correct option and four distractors. The MCQs were constructed by subject experts in accordance with the approved curriculum and examination blueprint to assess various cognitive levels.

After completion of all assesments and compilation of results, students' responses were anonymized, tabulated, and ranked according to their total examination scores. For the purpose of calculating the discrimination index, students were divided into upper and lower performance groups using the extreme group method. The top 27% of scorers constituted the upper group, while the bottom 27% formed the lower group, as recommended in classical test theory to maximize discrimination power. The middle 46% of students were excluded from discrimination analysis.

Each MCQ was evaluated using standard item analysis parameters, including difficulty index, discrimination index, and distractor efficiency. The difficulty index was calculated as the proportion of students answering an item correctly. The discrimination index was determined by comparing the proportion of correct responses between the upper and lower groups. Distractor efficiency was assessed by identifying non-functional distractors, defined as incorrect options selected by fewer than 5% of examinees. All analyses were performed solely for academic quality improvement and did not influence students' examination results.

.

Operational Definitions:

- **Difficulty Index :** Measures the easiness or difficulty of question items. The difficulty index was calculated as the proportion of students who answered an item correctly and categorized as difficult (<0.30), acceptable (0.30–0.70), or easy (>0.70) [21].
- **Discrimination Index:** Measures the ability

of question items to distinguish between high and low scorers. The discrimination index was determined by comparing the performance of high- and low-performing groups, with values ≥0.40 considered excellent, 0.30–0.39 good, 0.20–0.29 acceptable, and <0.20 poor[21].

- **Distractor Efficiency:** Measures the functionality of distractors. It was assessed by identifying non-functional distractors .A distractor is non-functional if chosen by <5% of student, items without non-functional distractors were considered to have optimal distractor efficiency[21].

The analysis was conducted solely for academic quality improvement and did not influence students' examination results.

Ethical considerations

Ethical approval was obtained from the Institutional Review Board of Dow university Of Health sciences

STATISTICAL ANALYSIS

Data obtained from students' response sheets were entered into Microsoft Excel and analyzed using SPSS version 22.0. Descriptive statistics were used to summarize the overall performance of the MCQs. For each item, the difficulty index (DI value), discrimination index (D-value), and distractor efficiency (DE) were calculated using standard formulas. MCQs were categorized based on predefined cutoff values for difficulty and discrimination indices to identify well-performing, acceptable, and poorly performing items. Distractor efficiency was determined by identifying non-functional distractors, defined as options selected by fewer than 5% of examinees.Frequencies and percentages were used to present the distribution of MCQs across different difficulty levels, discrimination categories, and distractor efficiency grades. All statistical analyses were performed at a descriptive level, **frequencies and percentages** were calculated for MCQs across different difficulty, discrimination, and distractor efficiency categories.

RESULTS

A total of 100 multiple-choice questions (MCQs) from a formative assesment of 3rd Professional year MBBS were analyzed based on responses obtained from (number) undergraduate medical students. Item analysis was performed using standard psychometric indices, including difficulty index, discrimination index, and distractor efficiency, to assess the overall quality of the examination items.

Analysis of the difficulty index showed that most MCQs were of acceptable difficulty. As presented in Table 1, 62% (n = 62) of the items fell within the acceptable difficulty range (0.30–0.70), indicating an appropriate balance between easy and challenging questions.

However, 23% (n = 23) of MCQs were classified as easy (> 0.70), suggesting that these items may have limited ability to assess higher cognitive levels. In contrast, 15% (n = 15) of items were categorized as difficult (< 0.30), possibly reflecting complex content or ambiguity in item construction.

Table 1 showed the distribution of MCQs According to Difficulty Index (n = 100)

Difficulty Level	Ranges	No. of MCQs (n)	Percentage (%)
Difficult	< 0.30	15	15.0
Acceptable	0.30–0.70	62	62.0
Easy	> 0.70	23	23.0

With regard to the discrimination index, the majority of MCQs demonstrated satisfactory discriminatory power. As shown in Table 2, 28% (n = 28) of items exhibited excellent discrimination (D ≥ 0.40), while 32% (n = 32) showed good discrimination (D = 0.30–0.39). An additional 21% (n = 21) of items demonstrated acceptable discrimination (D = 0.20–0.29).

Table 2 showed the distribution of MCQs According to Discrimination Index (n = 100)

Discrimination Category	Ranges	No. of MCQs (n)	Percentage (%)
Excellent	≥ 0.40	28	28.0
Good	0.30–0.36	32	32.0
Acceptable	0.20–0.29	21	21.0
Poor	< 0.20	19	19.0

However, 19% (n = 19) of MCQs had poor discrimination (D < 0.20), indicating limited ability to differentiate between high- and low-performing students and highlighting the need for revision or elimination of these items.

Evaluation of distractor efficiency revealed varying levels of functionality among incorrect options.

Table 3 showed the distractor Efficiency of MCQs (n = 100)

Non-Functional Distractors	Distractor Efficiency (%)	Number of MCQs	Percentage (%)
0	100%	45	45.0
1	66.6%	30	30.0
2	33.3%	17	17.0
≥3	0%	8	8.0

As illustrated in Table 3, 45% (n = 45) of MCQs demonstrated 100% distractor efficiency, with all distractors functioning effectively. 30% (n = 30) of items contained one non-functional distractor, while 17% (n = 17) had two non-functional distractors. A smaller proportion, 8% (n = 8), had three or more non-functional distractors, indicating poor distractor design.

Table 4 showed the Mean ± Standard Deviation of Item Analysis Parameters Derived from Distribution Ranges (n = 100 MCQs)

Item Analysis Parameter	Range Observed	Mean ± SD
Difficulty Index (P)	<0.30 – >0.70	**0.55 ± 0.18**
Discrimination Index (D)	<0.20 – ≥0.40	**0.31 ± 0.14**
Distractor Efficiency (%)	0 – 100	**71.2 ± 23.1**

Overall, the item analysis demonstrated that a substantial proportion of MCQs possessed acceptable difficulty levels, good discriminatory power, and effective distractor functioning.

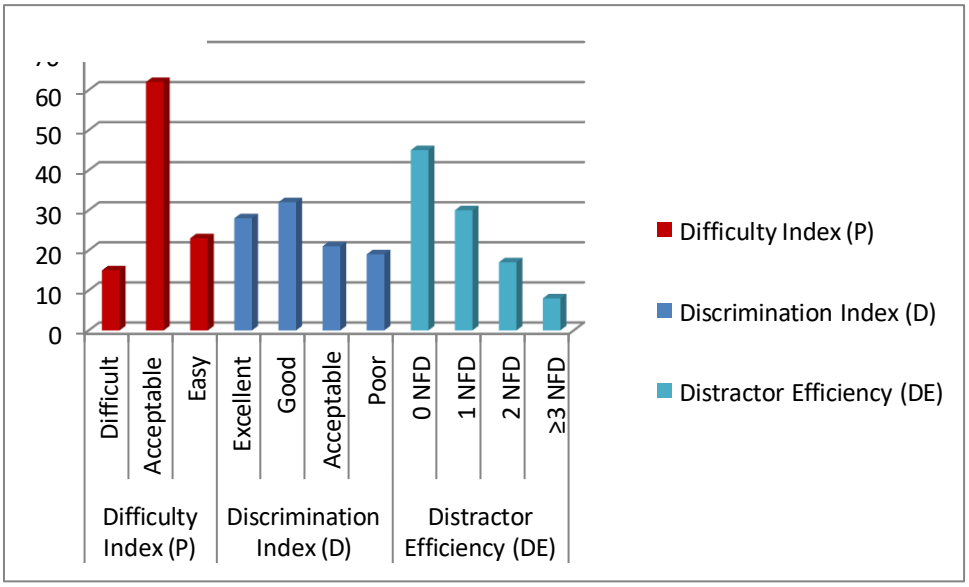

Figure 1 showed the Difficulty Index, discrimination Index and distractor Efficiency of MCQs

Nevertheless, nearly one-fifth of the items showed poor discrimination and low distractor efficiency, underscoring the need for systematic review and revision. These findings reinforce the importance of routine item analysis as a quality assurance tool for improving the psychometric robustness of MCQs used in formative medical assessments.

DISCUSSION

This study highlighted the importance of item analysis in improving the quality of multiple-choice questions (MCQs) in medical education. The findings demonstrated that most items fell within the acceptable range of difficulty, while a smaller proportion were either too easy or too difficult[10]. This distribution is desirable because moderately difficult questions are known to enhance discrimination and better assess students' conceptual understanding rather than rote memorization. Similar findings have been reported in previous medical education research, where the majority of well-constructed MCQs typically show a difficulty index between 30–70%, indicating optimal cognitive challenge[11,12].

The discrimination index results in the present study showed that several items effectively differentiated between high- and low-performing students, while

some exhibited poor discriminatory power. Studies conducted in various undergraduate medical settings have consistently emphasized that items with higher discrimination indices contribute significantly to examination validity[13].A research in basic medical sciences assessments has demonstrated that items with discrimination indices ≥0.30 are strong indicators of higher-order cognitive testing. Conversely, items with low or negative discrimination often reflect ambiguous wording, flawed key selection, or testing of trivial content, necessitating revision or removal[14].

Distractor efficiency analysis further revealed that a number of MCQs contained non-functional distractors. This observation aligns with previous literature suggesting that many MCQs suffer from poorly constructed distractors that fail to attract low-performing students[15,16]. Educational studies have shown that non-functional distractors reduce item quality and inflate guessing probability. In fact, several researchers have argued that three well-designed options may sometimes be more effective than four poorly constructed ones. Therefore, careful review and revision of distractors are essential to enhance assessment reliability[17].

The overall findings of this study are consistent with previous investigations in medical education, which confirm that routine post-examination item analysis significantly improves assessment standards[18]. Studies across disciplines such as physiology, pathology, and pharmacology have demonstrated that repeated cycles of item analysis lead to the development of high-quality question banks, improved internal consistency (Cronbach's alpha), and better alignment with learning outcomes in competency-based curricula.

Furthermore, in the context of competency-based medical education (CBME), assessment must not only measure knowledge recall but also evaluate application and clinical reasoning. Item analysis serves as an evidence-based feedback mechanism for faculty, helping identify whether MCQs are aligned with higher cognitive levels (Bloom's taxonomy) and clinical relevance. Regular implementation of item analysis also promotes faculty development in assessment literacy and strengthens institutional quality assurance processes[19,20].

LIMITATIONS

This study has several limitations that should be acknowledged. First, the analysis was conducted within a single institution, which may limit the generalizability of the findings to other medical colleges with different curricular structures or assessment policies. Second, the study relied primarily on classical test theory indices (difficulty index, discrimination index, and distractor efficiency), which, although widely used, provide limited insight into latent trait measurement compared to more advanced psychometric models such as Item Response Theory.

Additionally, the analysis was confined to quantitative metrics without incorporating qualitative review of item content, cognitive level classification, or faculty perspectives on item construction. The absence of differential item functioning (DIF) analysis also limits conclusions regarding potential subgroup bias. Furthermore, the cross-sectional nature of the study does not permit evaluation of item stability across multiple examination cycles…..u can add this under heading of limitations

CONCLUSION

The study highlighted that the Effective assessment is central to maintaining educational standards and ensuring competency in undergraduate medical education. This study demonstrates that systematic item analysis of multiple-choice questions—including evaluation of difficulty index, discrimination index, and distractor efficiency—provides valuable insights into the quality and reliability of examination items. Identifying non-performing items and understanding their psychometric properties allows educators to refine question banks, improve test fairness, and enhance student learning outcomes.

RECOMMENDATIONS

Multi-institutional studies incorporating differential item functioning (DIF) analysis are needed to ensure fairness, generalizability, and defensibility in high-stakes undergraduate medical examinations.

Interventional studies evaluating the impact of structured faculty development programs on MCQ quality would provide evidence for sustainable improvement strategies. Additionally, emerging research on artificial intelligence-assisted item review could enhance identification of flawed stems and non-functional distractors.

REFRENCES

1. Rezigalla, A. A., A. M. E. Eleragi, and M. Ishag. "Comparison between Students' Perception toward an Examination and Item Analysis, Reliability and Validity of the Examination." *Sudan Journal of Medical Sciences*, vol. 15, no. 2, 2020, pp. 114–123.
2. Rezigalla, A. A. "Item Analysis: Concept and Application." *Medical Education for the 21st Century*, edited by M. S. Firstenberg and S. P. Stawicki, London, 2022, pp. 1–16.
3. Awan, S. L., et al. "Types of Item Writing Flaws in Multiple Choice Questions in Medical Education and Their Effect on Passing Rate of Students." *Pakistan Journal of Physiology*, vol. 19, no. 4, 2023, pp. 49–52.
4. Sugianto, A. "Item Analysis of English Summative

Test: EFL Teacher-Made Test." *Indonesian EFL Journal*, vol. 1, no. 1, 2021, pp. 35–54.
5. Bhat, S. K., and K. H. L. Prasad. "Item Analysis and Optimizing Multiple-Choice Questions for a Viable Question Bank in Ophthalmology: A Cross-Sectional Study." *Indian Journal of Ophthalmology*, vol. 69, no. 2, 2021, pp. 343–346.
6. Shad, N. "Factors Affecting Difficulty & Discrimination Indices." *Medical Education*, 2016.
7. Obon, A. M., and K. A. M. Rey. "Analysis of Multiple-Choice Questions (MCQs): Item and Test Statistics from the 2nd Year Nursing Qualifying Exam in a University in Cavite, Philippines." *11th International Scholars Conference*, vol. 7, no. 1, 2019, pp. 499–511. doi:10.35974/isc.v7i1.1128.
8. Ramzan, M., et al. "Item Analysis of Multiple-Choice Questions at the Department of Community Medicine, Wah Medical College, Pakistan." *Life Sciences*, vol. 1, no. 2, 2020, pp. 60–63. doi:10.37185/LnS.1.1.53.
9. Sajjad, M., S. Iltaf, and R. A. Khan. "Nonfunctional Distractor Analysis: An Indicator for Quality of Multiple-Choice Questions." *Pakistan Journal of Medical Sciences*, vol. 36, no. 5, 2020, pp. 982–986. doi:10.12669/pjms.36.5.2439.
10. Aziz, A., et al. "Exploring the Experiences of Content Experts with Item Vetting during Item Bank Development." *Pakistan Journal of Medical Sciences*, vol. 40, no. 6, 2024, pp. 1241–1246.
11. Rezigalla, A. A., et al. "Item Analysis: The Impact of Distractor Efficiency on the Difficulty Index and Discrimination Power of Multiple-Choice Items." *BMC Medical Education*, vol. 24, no. 1, 2024. doi:10.1186/s12909-024-05433-y.
12. Alemu, A. T., et al. "Title Not Specified." *International Journal of Scientific Reports*, vol. 10, no. 6, 2024, pp. 195–199.
13. Tavakol, M., and R. Dennick. "Psychometric Evaluation of a Knowledge-Based Examination Using Rasch Analysis: An Illustrative Guide: AMEE Guide No. 72." *Medical Teacher*, vol. 35, no. 1, 2013, pp. e838–e848.
14. Patel, R. M. "Use of Item Analysis to Improve Quality of Multiple Choice Questions in II MBBS." *Journal of Education and Health Sciences*, vol. 4, no. 1, 2017, pp. 22–29.
15. Kumar, D., et al. "Item Analysis of Multiple Choice Questions: A Quality Assurance Test for an Assessment Tool." *Medical Journal Armed Forces India*, vol. 77, no. 1, 2021, pp. S85–S89.
16. Mehta, G., and V. Mokhasi. "Item Analysis of Multiple Choice Questions—An Assessment of the Assessment Tool." *International Journal of Health Sciences Research*, vol. 4, no. 7, 2014, pp. 197–202.
17. Tarrant, M., and J. Ware. "A Framework for Improving the Quality of Multiple-Choice Assessments." *Nurse Educator*, vol. 37, no. 3, 2012, pp. 98–104.
18. Velou, M. S., and E. Ahila. "Refine the Multiple-Choice Questions Tool with Item Analysis." *International Archives of Integrated Medicine*, vol. 7, no. 8, 2020, pp. 80–85.
19. Aljehani, D. K., et al. "Relationship of Text Length of Multiple-Choice Questions on Item Psychometric Properties—A Retrospective Study." *Saudi Journal of Health Sciences*, vol. 9, no. 2, 2020, pp. 84–87. doi:10.1159/000517662.
20. Uddin, I., et al. "Item Analysis of Multiple-Choice Questions in Pharmacology." *Journal of Saidu Medical College Swat*, vol. 10, no. 2, 2020, pp. 128–131.

