# Enhancing Assessment Quality in Optometry: Item Analysis of Core Diploma Examinations in Saudi Arabia

Nawal Omar Hawsawi[1], Omnia Adnan Ghandurah[2]

[1] Department of optometry, applied college, University of Bisha. ORCID ID: 0009-0004-4827-4438
[2] Department of optometry, King Abdulaziz Medical City, Ministry of National Guard Health Affairs. ORCID ID: 0000-0001-7911-0234

**ABSTRACT**:
To evaluate the psychometric quality of multiple-choice question (MCQ) examinations across four core courses in a Diploma in Optometry program at the University of Bisha, Saudi Arabia, with a focus on item difficulty, discrimination, distractor efficiency, and test reliability.
**Methods:** This retrospective descriptive study analyzed item performance data from final MCQ exams in four core optometry courses (Geometric Optics and Refraction, Ocular Anatomy and Physiology, Clinical Optometry Practice, and Ocular Diseases). For each item, we calculated the difficulty index (P-value), point-biserial discrimination index, and number of non-functional distractors (NFDs; distractors chosen by <5% of students). Items were categorized by difficulty (easy, acceptable, difficult) and discrimination (excellent, good, acceptable, poor) using established cut-offs. Distractor efficiency was classified according to the number of NFDs per item (0–3). Internal consistency reliability for each exam was estimated using Cronbach's alpha (Kuder–Richardson 20). Descriptive statistics summarized item metrics, and Pearson correlation analysis examined the relationship between item difficulty and discrimination.
**Results:** A total of 200 MCQs from four exams (40–60 items per course; 28–30 students per exam) were analyzed. Mean difficulty across exams ranged from 55.3% to 75.2%, with 14% of items classified as easy ($P \geq 0.78$), 10% as difficult ($P < 0.25$), and 76% as acceptable. Mean discrimination indices ranged from 0.22 to 0.35, and exam reliability was acceptable to good (Cronbach's $\alpha = 0.76$–$0.89$). Overall, 29% of items showed excellent discrimination (≥0.35), 23% good, 18.5% acceptable, and 29.5% poor, including ~1.5% with negative discrimination. Regarding distractor efficiency, 34% of items had no NFDs, 35% had one NFD, 21.5% had two NFDs, and 9.5% had three NFDs. Combining indices, 55% of items were classified as good, 35% as needing revision, and 10% as poor. A significant negative correlation was observed between item difficulty and discrimination ($r \approx -0.32$), with moderately difficult items demonstrating the highest discrimination.
**Conclusion:** The analyzed exams demonstrated generally acceptable psychometric properties, with most items falling within recommended ranges for difficulty and discrimination and all tests showing good internal consistency. However, a substantial proportion of items contained multiple non-functional distractors or suboptimal discrimination, particularly in the basic science course. Routine item analysis and systematic revision of flawed items are recommended to enhance MCQ quality, support fair and reliable assessment, and strengthen the optometry diploma curriculum.
**Keywords:** Item analysis; Difficulty index; Discrimination index; Distractor efficiency; Optometry; Multiple-choice exam.

## INTRODUCTION

High-quality assessment is essential in health professions education to ensure that graduates meet competency standards and learning outcomes. Multiple-choice questions (MCQs) are among the most

widely used assessment tools because they allow broad sampling of content, objective scoring, and efficient administration [1–3]. When carefully designed, MCQs can assess not only factual recall but also higher-order cognitive processes such as application and interpretation [1,3].

However, poorly constructed items—such as those with implausible distractors or ambiguous wording—can undermine the validity and reliability of examinations [3–5]. Item analysis is a post-examination procedure that uses quantitative indices to evaluate the performance of individual items and the test as a whole [2,4,6]. Commonly used indices include the difficulty index (P-value), which reflects the proportion of students answering correctly; the discrimination index, which reflects how well an item differentiates high- and low-performing students; and distractor efficiency, which evaluates the functioning of incorrect options [2–6,8].

Difficulty and discrimination are central to classical test theory. Moderately difficult items (often around 40–70% correct) typically exhibit the highest discrimination, while items that are extremely easy or extremely difficult tend to discriminate poorly [2,4,8,9]. Distractor analysis adds another layer of insight: non-functional distractors (NFDs), chosen by very few students (e.g., <5%), contribute little to item quality and may artificially increase the apparent easiness of the question [3–5]. Several studies have shown that a substantial proportion of teacher-constructed MCQs contain NFDs, and that items with more functional distractors tend to be more discriminating [3–5].

In medical, dental, and nursing education, numerous studies have used item analysis to evaluate and improve MCQ examinations [2,4–9]. These studies have highlighted common issues such as extreme difficulty levels, low discrimination indices, and a high frequency of NFDs. Psychometric analyses of high-stakes licensing examinations across health professions have also shown systematic relationships between item difficulty, discrimination, and reliability [1]. However, much of this work has focused on degree-level programs (e.g., medicine, nursing) and licensing exams rather than diploma-level programs.

Optometry education, particularly at the diploma level, has received less attention in the item analysis literature. Diploma optometrists often serve as frontline eye-care providers in primary care and community settings, making the quality of their assessments directly relevant to patient safety. In Saudi Arabia, psychometric analysis of MCQs has been reported in some health programs [7], but there is limited published evidence on the quality of examinations in optometry programs, especially in two-year diplomas.

The Diploma in Optometry program at the University of Bisha includes core courses covering optics, ocular anatomy and physiology, clinical optometry, and ocular diseases. Final course examinations largely rely on MCQs. Despite their frequent use, these examinations had not previously undergone systematic item analysis.

Therefore, this study aimed to conduct a comprehensive item analysis of MCQ-based final examinations across four core courses in the Diploma in Optometry program at the University of Bisha. The specific objectives were to:

1. Describe the distribution of item difficulty, discrimination, and distractor efficiency across four core courses.
2. Estimate the internal consistency reliability of each examination.
3. Classify items into quality categories (good, needs revision, poor) based on combined indices and identify priorities for item revision and test improvement.

# METHODS

## Study Design and Setting

This retrospective descriptive study was conducted at the Diploma in Optometry program, Applied College, University of Bisha, Saudi Arabia. The program is a two-year diploma designed to prepare mid-level optometry professionals. The study focused on final MCQ examinations from four core courses administered during the 2024 academic year.

## Courses and Participants

Four courses were included:

1. **Geometric Optics and Refraction**
2. **Ocular Anatomy and Physiology**
3. **Clinical Optometry Practice**
4. **Ocular Diseases**

These courses represent key domains of the diploma curriculum (optics, basic ocular science, clinical practice, and ocular pathology). All students enrolled in each course who sat for the final examination were included. Class sizes ranged from 28 to 30 students per course.

## Examination Format

Each final examination consisted exclusively of single-best-answer MCQs with four options (one correct answer and three distractors). The number of items per exam ranged from 40 to 60, yielding a total of 200 items across the four courses (40, 50, 60, and 50 items, respectively). All items carried equal weight, and there was no negative marking.

Examinations were administered under standard proctored conditions. Student responses were recorded using optical mark reader (OMR) sheets and processed by the university's examination system. Item-level response data (number and percentage of students selecting each option, including the correct answer) and total test scores were exported in de-identified format for analysis.

## Item Analysis Indices

For each item, we calculated the following indices based on classical test theory [2,4,6,8,9].

## Difficulty Index (P-value)

The difficulty index (P) was defined as the proportion of students who answered the item correctly:

*P = (Number of students answering correctly / Total number of students).*

Higher P-values indicate easier items. Items were categorized as [2,4,8]:

- **Easy:** $P \geq 0.78$
- **Acceptable (moderate):** $0.25 \leq P < 0.78$
- **Difficult:** $P < 0.25$

Mean difficulty indices were computed for each exam, and the distribution of items across difficulty categories was summarized.

## Discrimination Index (Point-Biserial Correlation)

Item discrimination was calculated using the point-biserial correlation coefficient (r_pb), which quantifies the relationship between each item's dichotomous score (1 = correct, 0 = incorrect) and the corrected total test score (total score minus the item score) to avoid part-whole inflation.

Items were categorized as [2,4,8,9]:

- **Excellent:** $r_pb \geq 0.35$
- **Good:** 0.25–0.34
- **Acceptable (marginal):** 0.20–0.24
- **Poor:** $r_pb \leq 0.19$

Items with negative discrimination were flagged as potentially flawed (e.g., ambiguous wording, miskeyed answers, or misaligned content).

**Distractor Efficiency (Non-functional Distractors)**

Each item had three distractors. A distractor selected by fewer than 5% of students was considered a non-functional distractor (NFD) [2,3,5]. For each item, we counted the number of NFDs (0–3) and classified distractor efficiency as follows:

- **0 NFDs:** all distractors functional (excellent efficiency)
- **1 NFD:** good efficiency
- **2 NFDs:** low efficiency
- **3 NFDs:** very poor efficiency

The distribution of items by NFD count was summarized overall and by course.

**Test Reliability**

Internal consistency reliability for each exam was estimated using Cronbach's alpha, which is equivalent to the Kuder–Richardson Formula 20 (KR-20) for dichotomous items [1,6]. Reliability coefficients were interpreted as:

- ≥0.70: acceptable
- ≥0.80: good
- ≥0.90: excellent

**Overall Item Quality Classification**

Based on combined item statistics (difficulty, discrimination, and distractor performance), each item was classified into one of three overall quality categories [2–5,8]:

- **Good:** difficulty in the acceptable range; discrimination ≥0.20; and ≤1 NFD.
- **Needs revision:** suboptimal statistics (e.g., very easy or difficult, discrimination 0.10–0.19 and/or 2 NFDs) but potentially improvable.
- **Poor:** negative or near-zero discrimination, extreme difficulty ($P > 0.95$ or $P < 0.10$), and/or 3 NFDs.

These categories were used to estimate the proportion of items suitable for retention, revision, or replacement.

**Data Analysis**

Data were entered into Microsoft Excel and analyzed using SPSS (version 27; IBM Corp., Armonk, NY, USA). Descriptive statistics (means, standard deviations, frequencies, and percentages) summarized item and test-level indices. Pearson correlation coefficients were calculated to examine the relationship between item difficulty (P) and discrimination (r_pb) across all items. Statistical significance was set at $P < 0.05$.

## RESULTS

**Exam characteristics**

A total of 200 MCQ items from four core course examinations were analyzed. Table 1 summarizes exam characteristics and psychometric indices by course.

**Table 1. Exam characteristics and psychometric indices by course**

| Course | Students (n) | Items (n) | Mean difficulty % (SD) | Mean discrimination (SD) | Reliability (Cronbach's α) |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Geometric Optics and Refraction | 30 | 40 | 65.0 (18.2) | 0.30 (0.15) | 0.82 |
| Ocular Anatomy and Physiology | 30 | 50 | 75.2 (15.4) | 0.22 (0.12) | 0.76 |
| Clinical Optometry Practice | 28 | 60 | 55.3 (20.5) | 0.35 (0.18) | 0.89 |
| Ocular Diseases | 28 | 50 | 62.1 (17.0) | 0.28 (0.16) | 0.80 |
| **Overall (unweighted mean)** | – | 200 | 64.7 (19.1) | 0.28 (0.19) | – |

All examinations demonstrated acceptable to good internal consistency, with Cronbach's alpha ranging from 0.76 to 0.89.

**Difficulty index**

Across all 200 items, the overall mean difficulty index was 64.7% (SD 19.1%). Using the predefined cut-offs:

- **Easy (P ≥ 0.78):** 28 items (14%)
- **Acceptable/moderate (0.25 ≤ P < 0.78):** 152 items (76%)
- **Difficult (P < 0.25):** 20 items (10%)

Ocular Anatomy and Physiology was the easiest exam, with a mean difficulty of 75.2%. Clinical Optometry Practice was the most difficult exam, with a mean difficulty of 55.3%. The Geometric Optics and Refraction and Ocular Diseases exams showed relatively balanced difficulty distributions, with most items in the moderate range and few extremes.

**Discrimination index**

The overall mean item discrimination index was 0.28 (SD 0.19). Item-level discrimination was distributed as follows:

- **Excellent (r_pb ≥ 0.35):** 58 items (29%)
- **Good (0.25–0.34):** 46 items (23%)
- **Acceptable/marginal (0.20–0.24):** 37 items (18.5%)
- **Poor (r_pb ≤ 0.19):** 59 items (29.5%), including 3 items (1.5%) with negative discrimination

Clinical Optometry Practice exhibited the highest mean discrimination (0.35), whereas Ocular Anatomy and Physiology had the lowest mean discrimination (0.22). Items with very high or very low difficulty tend to show lower discrimination.

Across all items, there was a statistically significant negative correlation between difficulty and discrimination ($r \approx -0.32$, $P < 0.001$), indicating that easier items were generally less discriminating, whereas moderately difficult items tended to have higher discrimination, consistent with previous findings in health professions education [1,2,4,8].

**Distractor efficiency**

Table 2 summarizes the distribution of items according to the number of non-functional distractors (NFDs).

**Table 2. Number of non-functional distractors (NFDs) per item (N = 200)**

| NFDs per item | Items (n) | Percentage of total items (%) |
|---|---|---|
| 0 NFDs | 68 | 34.0 |
| 1 NFD | 70 | 35.0 |
| 2 NFDs | 43 | 21.5 |
| 3 NFDs | 19 | 9.5 |

Overall, 69.0% of items had ≤1 NFD, indicating good to excellent distractor efficiency. However, nearly one-third of items (31.0%) had two or three NFDs, and 9.5% of items had three NFDs, indicating that none of the distractors were functioning adequately. Items with three NFDs were typically very easy and had low discrimination.

**Overall item quality**

Based on combined criteria for difficulty, discrimination, and distractor performance, items were classified into three quality categories (good, needs revision, poor). Table 3 presents the overall item quality classification.

**Table 3. Overall item quality classification (N = 200)**

| Item quality category | Number of items (n) | Percentage of total items (%) |
|---|---|---|
| Good | 110 | 55.0 |
| Needs revision | 70 | 35.0 |
| Poor | 20 | 10.0 |
| **Total** | 200 | 100.0 |

Thus, just over half of the items (55.0%) were considered suitable for retention with minimal or no modification, whereas 35.0% required revision and 10.0% were classified as poor and candidates for major revision or replacement.

## DISCUSSION

This study provides a detailed classical item analysis of MCQ-based final examinations across four core courses in a diploma-level optometry program. Overall, the examinations demonstrated moderate difficulty, acceptable discrimination, and good internal consistency, which aligns with recommended psychometric standards for educational assessments in health professions [2,4,6–9]. At the same time, the analysis highlighted specific areas for improvement, particularly regarding distractor efficiency and the proportion of items with low discrimination.

**Difficulty and exam balance**

The mean difficulty indices for the four exams (55.3–75.2%) indicate that the tests were generally of moderate difficulty. Most items (76.0%) fell within the acceptable difficulty range, with smaller proportions classified as easy (14.0%) or difficult (10.0%). This distribution is broadly consistent with recommendations that the majority of MCQs should be of moderate difficulty, with a smaller subset of easier and more challenging questions to differentiate performance levels [2,4,8].

The Ocular Anatomy and Physiology exam was notably easier than the others, with the highest mean difficulty index and the highest proportion of easy items. This pattern suggests a predominance of straightforward, recall-level questions. While such items can confirm basic factual knowledge, an overrepresentation of very easy items may limit the exam's ability to distinguish between higher- and

lower-performing students and may contribute to lower discrimination and reliability [4,8]. In contrast, the Clinical Optometry Practice exam was more challenging and contained a somewhat higher proportion of difficult items, reflecting its emphasis on application and clinical reasoning.

Overall, the difficulty profile observed in this study is comparable to item analyses reported in medical and allied health education, which often report mean item difficulty in the 60–70% range [2,4,6–9]. For a diploma-level optometry program, the current balance appears acceptable; however, the basic science components, particularly anatomy and physiology, could benefit from more moderately challenging items that target deeper understanding rather than simple recall.

**Discrimination and reliability**

The mean discrimination indices across exams (0.22–0.35) and the substantial proportion of items with good or excellent discrimination (52.0%) indicate that many MCQs effectively differentiated between higher- and lower-achieving students. The strongest discrimination was observed in the Clinical Optometry Practice exam, which also had the highest reliability ($\alpha = 0.89$). This pattern is consistent with evidence that examinations containing more moderately difficult, scenario-based items often achieve stronger discrimination and internal consistency [1,2,4,7–9].

The statistically significant negative correlation between difficulty and discrimination ($r \approx -0.32$) is in line with classical test theory and prior psychometric analyses of health professions examinations [1,2,4]. As expected, extremely easy or extremely difficult items tended to have lower discrimination, whereas items in the moderate difficulty range exhibited higher discrimination. This finding underscores the importance of deliberate exam blueprinting to avoid an overconcentration of very easy or very hard items.

All four exams exhibited acceptable to good reliability (Cronbach's $\alpha = 0.76–0.89$), which is encouraging given the relatively small cohort sizes. These values are comparable to, or higher than, reliability estimates reported in similar educational settings and suggest that the exams provide reasonably consistent measures of student performance [6,7].

**Distractor performance**

Distractor analysis revealed that, while a majority of items had adequate distractor efficiency (0–1 non-functional distractor [NFD]), nearly one-third had low efficiency (2–3 NFDs). Approximately 10% of items had three NFDs, meaning that none of the distractors attracted at least 5% of students. These items were typically very easy and had low discrimination, consistent with the literature showing that non-functional distractors reduce item quality and provide minimal additional information about student ability [3–5].

The proportion of items with all distractors functioning (34.0%) compares favorably with some previous studies of teacher-constructed MCQs in nursing and medical education, which have reported lower rates of fully functional distractors [3–5]. Nonetheless, the sizable proportion of items with multiple NFDs indicates that distractor construction remains an important area for improvement, especially in basic science courses such as anatomy and physiology.

Evidence suggests that replacing or removing non-functional distractors can improve item performance without compromising exam validity, and that three-option MCQs may perform as well as four-option items when it is difficult to generate a fourth high-quality distractor [3–5,8]. In the present context, targeted revision of distractors, informed by common misconceptions observed in student responses, could meaningfully enhance exam quality.

These findings are consistent with established item-writing guidelines, which emphasize that distractors should be plausible, homogeneous, and free from obvious cues, and warn that options rarely selected by examinees add little psychometric value to an item [10]. In line with Rodriguez's meta-analysis showing that three-option MCQs can perform as well as or better than four-option

items when distractors are well constructed [11], our results suggest that instructors should prioritize the quality rather than the number of distractors when developing or revising test items.

**Comparison with previous studies**

The findings of this study are broadly in line with item analyses conducted in other health professions education settings, which typically report moderate item difficulty, acceptable discrimination, and a subset of items requiring revision [2,4,6–9]. Psychometric analyses of national licensing examinations have also documented inverse relationships between item difficulty and discrimination and emphasized the value of routine item analysis as part of quality assurance [1].

Our results complement previous work by providing data from a diploma-level optometry program in Saudi Arabia, a context that has been underrepresented in the literature. The study demonstrates that even with relatively small cohorts, systematic item analysis yields actionable information that can guide targeted improvements in question banks and exam design.

Our results also parallel findings from other item-analysis studies in medical education, which have reported comparable patterns of moderate mean difficulty, variable discrimination indices, and a sizeable subset of items requiring revision or replacement [12,13].

**Implications for assessment in health professions education**

For diploma-level optometry education and health professions education more broadly, robust, data-informed assessment is essential to ensure that graduates are adequately prepared for clinical practice. The results of this study suggest that the core examinations in the University of Bisha's Diploma in Optometry program provide a reasonable foundation but also highlight clear opportunities for enhancing item and exam quality.

Embedding routine item analysis into the assessment cycle, providing faculty development in MCQ construction, and establishing formal criteria for item retention and revision can strengthen the reliability and validity of assessments and align them more closely with program learning outcomes. In the long term, such measures can support better preparation of graduates for external licensing examinations and clinical practice across a range of health professions.

Drawing on consolidated item-writing guidelines [10] and empirical evidence favoring fewer, but higher-quality distractors [11], embedding structured faculty development on MCQ construction could help reduce the proportion of low-performing items and strengthen exam validity. Furthermore, integrating routine item and test analysis procedures similar to those described by Gajjar et al. and Rao et al. [12,13] would support continuous quality improvement of assessments across courses and programs, beyond the optometry diploma alone.

## CONCLUSION

Final MCQ examinations across four core courses in a Diploma in Optometry program at the University of Bisha demonstrated generally acceptable psychometric quality, characterized by moderate item difficulty, satisfactory discrimination, and good internal consistency. Most items fell within the recommended range of difficulty, and more than half showed good overall performance when difficulty, discrimination, and distractor efficiency were considered together.

At the same time, the analysis revealed several areas that need improvement. A substantial proportion of items, particularly in basic science examinations, exhibited low discrimination and multiple non-functional distractors, limiting their ability to differentiate between higher- and lower-performing students. These findings underscore the importance of systematic, data-driven review of examination items.

**Recommendations**

Based on the findings of this item analysis, the following recommendations are proposed to enhance the quality of MCQ-based examinations in health professions and allied health education programs:

1. **Implement routine item analysis**

Conduct item analysis after each major examination and formally review item difficulty, discrimination, and distractor efficiency. Summarized results should be discussed at course, department, and/or program level as part of regular educational quality assurance.

2. **Revise or remove flawed items**

Identify items with negative or very low discrimination, extreme difficulty (very easy or very difficult), or three non-functional distractors. For these items, revise stems, answer keys, and distractors as appropriate. Items that continue to perform poorly after revision should be removed from the active question bank.

3. **Improve distractor quality**

Provide focused faculty development on constructing plausible distractors that reflect realistic misconceptions or common errors. Instructors should avoid obviously incorrect, irrelevant, or heterogeneous options. When it is difficult to generate a fourth high-quality distractor, three-option MCQs may be considered, in line with evidence that well-constructed three-option items can perform as well as four-option items.

4. **Balance exam difficulty through blueprinting**

Use explicit exam blueprints that specify the desired mix of easy, moderate, and difficult items for each assessment, with the majority of items in the moderate range. Blueprinting should also ensure appropriate coverage of content domains and cognitive levels. This approach can help optimize discrimination and reliability and prevent overconcentration of very easy or very hard items.

5. **Increase cognitive level where appropriate**

Convert very easy, purely recall-based questions into items that require application, interpretation, or clinical reasoning, especially in foundational science courses. This shift can enhance item discrimination and align assessments more closely with the competencies expected of graduates in health professions and allied health programs.

6. **Develop and maintain an annotated item bank**

Establish a centralized, annotated item bank that records, for each question, its content area, cognitive level, and performance indices (difficulty, discrimination, and distractor statistics) across cohorts. Such an item bank supports evidence-based selection of items for future examinations, facilitates continuous monitoring of assessment quality, and reduces reliance on ad hoc item construction.

**Limitations**

These recommendations, while derived from an item analysis in a diploma-level optometry program, can be applied to a broad range of health professions and allied health education settings seeking to strengthen the psychometric quality of their MCQ-based assessments.

This study has several limitations that should be considered when interpreting the findings. First, the analysis was conducted in a single diploma-level optometry program with relatively small cohort sizes (28–30 students per exam). Small samples may reduce the stability and precision of item statistics, particularly discrimination indices, and limit the generalizability of the results to other institutions or programs.

Second, the study was based on classical test theory; more advanced psychometric approaches such as item response theory were not employed. While classical indices (difficulty, discrimination, distractor efficiency, and reliability) are appropriate and widely used for practical exam review, they do not provide information on item functioning across different ability levels or allow for more sophisticated modeling of student performance.

Third, only MCQ-based written examinations were analyzed. Other important components of assessment in optometry education—such as practical examinations, OSCEs, clinical evaluations, and assignments—were not included. As a result, the study does not offer a complete picture of overall assessment quality at the program level.

Fourth, although problematic items were identified using quantitative indices, a detailed qualitative content review of each item (for example, examining stems, keys, and distractors for specific flaws) was beyond the scope of this report. Therefore, suggested causes for poor performance (e.g., ambiguity, misalignment with taught content, or superficial recall) remain inferential rather than confirmed.

Finally, comparisons with previous studies must be interpreted cautiously, as differences in curricular context, exam stakes, cut-off values, and student populations may influence item performance. Future research could address these limitations by analyzing multiple cohorts, incorporating qualitative item review, including additional assessment modalities, and applying more advanced psychometric models.

**Ethical Considerations**

This study involved secondary analysis of de-identified examination data and did not include any intervention, contact with students, or collection of personal or health information. In accordance with the institutional and national guidelines at the University of Bisha, it met the criteria for exemption from full Institutional Review Board (IRB) review, and the requirement for written informed consent was waived. Administrative permission to access and analyze anonymized item-analysis reports was obtained from the Applied College and the Department of Optometry.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. All data access and analytic activities were conducted as part of routine academic quality assurance within the University of Bisha.

**Data Availability**

The de-identified item analysis dataset generated and analyzed during the current study is available from the corresponding author on reasonable request, subject to institutional policies and examination security considerations. Item-level content is not publicly shared to protect the integrity of the question bank and maintain exam confidentiality.

**Contribution of Authors**

The Authors conceived and designed the study, coordinated data collection, performed the statistical analysis, and drafted and revised the manuscript. The authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

**Competing Interests**

The authors declare no competing interests. This study was conducted as part of an internal educational quality improvement initiative, and there were no financial or personal relationships that could have influenced the design, analysis, or reporting of the findings.

**References**

1. Kim, Y. H., Kim, B. H., Kim, J., Jung, B., & Bae, S. (2025). Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2024, Korea: A psychometric study. *Journal of Educational Evaluation for Health Professions*, *22*, 38.

2. Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, *24*(1), 445.

3. Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, *9*(1), 40.

4. Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2018). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Ann Pak Inst Med Sci.* 2017;13(4):310–15.

5. D'Sa, J., & Visbal-Dionaldo, M. (2017). Analysis of Multiple Choice Questions: Item Difficulty, Discrimination Index and Distractor Efficiency. *International Journal of Nursing Education*, *9*, 109.

6. Elgadal, A., & Mariod, A. (2021). Item Analysis of Multiple-choice Questions (MCQs): Assessment Tool For Quality Assurance Measures. *Sudan Journal of Medical Sciences*, *16*, 334–346.

7. Salih, K. E. M. A., Jibo, A., Ishaq, M., Khan, S., Mohammed, O. A., AL-Shahrani, A. M., & Abbas, M. (2020). Psychometric analysis of multiple-choice questions in an innovative curriculum in Kingdom of Saudi Arabia. *Journal of Family Medicine and Primary Care*, *9*(7), 3663–3668.

8. Namdeo, S. K., & Sahoo, B. (2016). Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. *International Journal of Research in Medical Sciences*, *4*(5), 1716–1719.

9. Zafar, U., Ali, Z., Khalid, A., Zaki, S., Naeem, F. J., & Chaudhry, M. A. (2022). Difficulty index, discrimination index, sensitivity and specificity of single best answer questions to assess medical students' performance in send-up examination at Lahore Medical & Dental College. *Pakistan Journal of Physiology*, *18*(3), 44–47.

10. Haladyna, T., Downing, S., & Rodriguez, M. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, *15*, 309–333.

11. Rodriguez, M. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice*, *24*, 3–13.

12. Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine : Official Publication of Indian Association of Preventive & Social Medicine*, *39*(1), 17–20.

13. Rao, C., Prasad HL, K., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, *2*, 201.