Bhutan Health Journal

ORIGINAL ARTICLE

OPEN ACCESS

https://doi.org/10.47811/bhj.207

Item analysis of pharmacology multiple-choice questions used in the semester-I examination for second-year undergraduate medical students at Khesar Gyalpo University of Medical Sciences, Bhutan: a cross-sectional study

Kipchu Tshering[1] , Gyem Dorji[2], Kanokwan Wetasin[3]

[1]Assistant Professor, Department of Pharmacology, Faculty of Undergraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan

[2]Associate Professor, Department of Anatomy, Faculty of Undergraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan.

[3]Chief of Healthcare Quality Development Department, Sakon Nakhon Hospital, Thailand

ABSTRACT

Background: As the Faculty of Undergraduate Medicine is a newly established medical school in Bhutan, evaluation of the quality of assessment tools is essential. This study aimed to evaluate the quality of Pharmacology MCQs used in semester-end summative examinations. **Methods:** MCQs used in the Semester I summative examination conducted in June 2025 for second-year MBBS students were analysed. A total of 50 MCQs were included in the study. Item analysis of 50 MCQs from the Semester-I Pharmacology examination administered to 24 second-year MBBS students was performed using difficulty index, discrimination index, distractor effectiveness, and Cronbach's alpha reliability analysis. An ethical waiver was obtained from the Institutional Review Board of the University. Statistical analysis was carried out using SPSS version 26. **Results**: Item analysis of the 50 MCQs showed that 60% of items had an acceptable (moderate) difficulty index, while 24% were categorized as very difficult and 16% as easy. Regarding the discrimination index, 64% of items demonstrated good to excellent discrimination, 20% showed fair discrimination, and 16% exhibited poor discrimination. Distractor analysis revealed that 67% of items had no non-functioning distractors, although a small proportion (3%) contained three or more non-functioning distractors. Reliability analysis demonstrated acceptable internal consistency, with a Cronbach's alpha value of 0.78 for the overall assessment. **Conclusion:** The majority of Pharmacology MCQs demonstrated acceptable difficulty and discrimination, with a Cronbach's alpha of 0.78. A proportion of items require revision, and routine item analysis is recommended as standard practice in summative assessment quality assurance.

Keywords: *Difficulty index; Discrimination index; Distractor analysis; Item analysis; Multiple-choice questions*

INTRODUCTION

Assessment is a cornerstone for undergraduate medical education, guiding both teaching and learning activities[1]. Multiple-choice questions (MCQs) remain one of the most widely used assessment tools due to their objectivity, reliability, and efficiency in evaluating a wide range of knowledge[2]. However, the effectiveness of MCQs in discriminating between high- and low-performing students and in validly assessing intended learning outcomes largely depends on their quality[2,3].

The undergraduate medical programme in Bhutan is in its early stage. The Faculty of Undergraduate Medicine was established in 2023 and is currently in its third year of operation, with the third cohort of MBBS students enrolled. The curriculum has been newly developed by our own faculty, who also lack prior experience in undergraduate teaching and assessment. As the sole institution in the country mandated to train medical doctors, it is imperative to rigorously assess the quality of multiple-choice questions (MCQs) used in module-end summative assessments. High-quality MCQs are essential to accurately evaluate students' performance and to reinforce effective teaching and learning methodologies[4]. Conversely, poorly constructed MCQs risk misjudging student competence and compromising educational outcomes[5].

According to the Bachelor of Medicine and Bachelor of Surgery (MBBS) curriculum of the Faculty of Undergraduate Medicine, 50 marks in Paper III of Module 3 are allocated to Basic Pharmacology, with the remaining marks assigned to structured essay questions (SEQs). Each MCQ consists of a stem followed by four options in a single best answer format. This examination structure informed the analytical approach adopted in the present study.

Therefore, this study aims to comprehensively evaluate the quality of MCQs used in semester end summative assessments of second-year undergraduate medical students

Corresponding author:
Kipchu Tshering
kipchu@kgumsb.edu.bt

in Bhutan. The objectives are to assess item quality using established psychometric indices, to evaluate alignment with published academic standards, and to identify reliable, high-quality MCQs suitable for inclusion in an institutional question bank[6]. The findings are expected to provide scientific evidence to support the validity, reliability, and acceptability of MCQs used in our newly established faculty.

METHODS

Study Design and Setting

This cross-sectional descriptive study was conducted at the Faculty of Undergraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan. The study utilized the MCQ results of the second year Semester I examination conducted in June 2025 for the Pharmacology paper.

Study Participants

A total of 24 second-year MBBS students appeared for the Semester I examination in June 2025 in the Pharmacology paper. As per the curriculum, Semester I covers four modules, namely the nervous system, the special senses, basic pharmacology and pathology. These modules are assessed through four papers conducted in June 2025 as the year II semester I end assessment examination.

Examination Structure

Each paper consisted of 50 marks of multiple-choice questions (MCQs) and 50 marks of structured essay questions (SEQs). For this study, only the 50 MCQs used in the Pharmacology paper were included. All MCQs were of the single best response type, comprising one stem with four options, including one correct answer and three distractors. Each question carried one mark, with a maximum possible score of 50 and there are no negative marking for the wrong answers. Only the answer script of the MCQs of 24 students were analyzed. On the answer scripts, students used only their unique index numbers and did not include their names which was a mandate in any examination to prevent identification of students. During data entry, additional unique codes were assigned to each student; therefore, the dataset was fully anonymized, ensuring the protection of privacy and personal identifiers.

Item Analysis

The results of all papers were ranked in descending order, from highest to lowest scores. The top 27% (n=7) of students were categorized as the high-score group. The 27% cut-off was applied following the method described by Shrinivas et al. (2023), which is established for use in small cohorts and provides adequate separation between extreme group[7]. Item analysis was conducted using the following indices:

1.Difficulty Index (DIF):

DIF represents the proportion of students who answered a question correctly. It was calculated using the formula (7): DIF=(H+L) / N, where H = number of students answering correctly in the high-score group, L = number of students answering correctly in the low-score group, and N = total number of students in both groups combined. The DIF ranges from 0 to 1; when multiplied by 100, the p-value represents the percentage of students who answered correctly (Table 1).

1: Difficulty Index used in the study

Range (%)	Interpretation	% Correct	Action
0 – 0.25	Very difficult	0–25%	Revise or discard
0.26 – 0.75	Right difficulty	26–75%	Retain
≥0.76	Easy	>76%	Revise or discard

2. Discrimination Index (DI):

Discrimination Index (DI) measures the ability of an item to differentiate between high- and low-performing students. It was calculated using the formula (7): DI =2(H − L)/N, where H = number of correct responses in the high-score group, L = number of correct responses in the low-score group, and N= total number of students in both groups combined (N=14). The denominator is 7, yielding a DI range from −1 to +1. Higher positive values indicate better discrimination (Table 2).

The following classification by Shrinivas et al. (2023) was applied consistently throughout this study:

Table 2: Discrimination Index Classification

DI Range	Interpretation	Quality	Action
≥0.50	Excellent	Excellent	Definitely retain
0.40 – 0.49	Good	Very usable	Retain
0.20 – 0.39	Fair	Usable	Revise
≤0.20	Poor	Poor	Discard

3. Distractor Effectiveness (DE):

Distractor Effectiveness (DE) assess the performance of distractors. A distractor chosen by fewer than 5% of the total student cohort (i.e. fewer than 1 or 2 out of 24 students, effectively 0 or 1 student) was classified as a non-functioning distractor (NFD); distractors selected by 5% or more were considered functional distractors (FD). Distractor effectiveness was assessed at the item level, based on the number of NFDs per item: 100% = no NFDs; 66.6% = 1 NFD; 33.3% = 2 NFDs; 0% = 3 or more NFDs.

4. Reliability

The reliability of the examination was assessed using Cronbach's alpha coefficient. Interpretation: ≥0.90 = excellent; 0.80–0.89 = good; 0.70–0.79 = acceptable; 0.60–0.69 = questionable; 0.50–0.59 = poor; <0.50 = unacceptable[7].

Data Analysis

Data were entered into Microsoft Excel 2010 and analysed using SPSS version 26. Quantitative variables (DIF and DI) are reported as mean ± standard deviation (SD) and as frequency and percentage by category. Qualitative variables are presented as frequency and percentage.

Ethical Considerations

Ethical waiver for the study was obtained from the Institutional Review Board (IRB) of Khesar Gyalpo University of Medical Sciences of Bhutan with Ref. No. IRB/Waiver-Exempt/PN/2026/001/ 179 dated 13th February 2026. Administrative and site clearance were obtained from the Faculty of Undergraduate Medicine and Khesar Gyalpo University of Medical Sciences of Bhutan, respectively, before accessing examination papers.

RESULTS

Twenty-four second year undergraduate medical students appeared for the examination in June 2025. All the MCQs from the Basic Pharmacology semester-end examination were analysed for item difficulty, discrimination index, distractor effectiveness, and reliability. There were 50 MCQs in total and every student attempted the questions. The mean difficulty index was 0.51 ± 0.18 and the mean discrimination index was 0.31 ± 0.14.

Analysis of the difficulty index revealed that 12 items (24%) were categorised as very difficult, 8 items (16%) as easy, and 30 items (60%) demonstrated an acceptable or moderate level of difficulty (Table 3).

Table 3: Classification of Pharmacology Multiple Choice Questions used in Semester-I Examination of Second-Year MBBS Students at Khesar Gyalpo University of Medical Sciences of Bhutan according to difficulty index (DIF) (n=50)

Difficulty Index (p)	Interpretation	Items (%)	Action
0 – 0.25	Difficult	12 (24)	Revise or discard
0.26 – 0.75	Right difficulty	30 (60)	Retain
≥0.76	Easy	8 (16)	Revise or discard

Regarding item discrimination, using the Shrinivas et al. classification (0.20–0.39 = fair; 0.40–0.49 = good; ≥0.50 = excellent): 10 items (20%) showed excellent discrimination, 22 items (44%) demonstrated good discrimination, 10 items (20%) showed fair discrimination, and 8 items (16%) exhibited poor discrimination (Table 4).

Table 4: Classification of Pharmacology Multiple Choice Questions used in Semester-I Examination of Second-Year MBBS Students at Khesar Gyalpo University of Medical Sciences of Bhutan according to discrimination index (DI) (n=50)

DI Range	Interpretation	Items (%)	Action
≤0.20	Poor	8 (16)	Discard
0.20 – 0.39	Fair	10 (20)	Revise
0.40 – 0.49	Good	22 (44)	Retain
≥0.50	Excellent	10 (20)	Definitely retain

With respect to distractor effectiveness, the analysis was performed at the item level (n = 50 items, with 3 distractors per item = 150 distractors total). At the item level, 33 (67%) items contained no non-functioning distractors,10 items (20%) had one NFD; 5 items (10%) had two NFDs; and 1 item (2%) (reported as 3%, rounded) had three or more NFDs. The distribution of distractor effectiveness is illustrated in Figure 1.

Reliability analysis of the 50 MCQs is reported separately in Table 5. The Cronbach's alpha coefficient was 0.78, indicating acceptable internal consistency for summative assessment purposes.

Table 5: Reliability of the Pharmacology MCQ paper (n = 50 items)

Measure	Value	Interpretation
Cronbach's alpha	0.78	Acceptable (0.70–0.79)
Number of items	50	

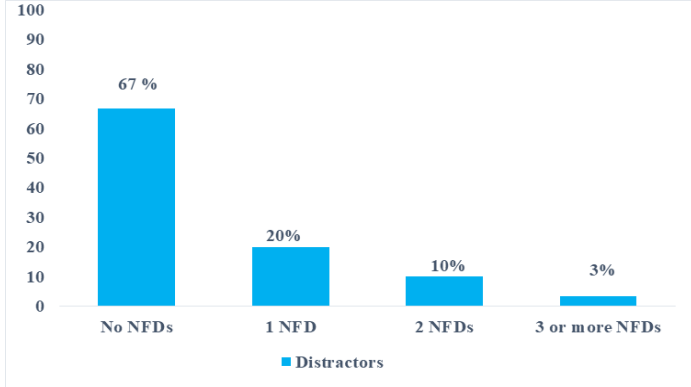


Figure 1: Distractor Effectiveness (DE) of Pharmacology Multiple Choice Questions used in Semester-I Examination of Second-Year MBBS Students at Khesar Gyalpo University of Medical Sciences of Bhutan(n-150).

DISCUSSION

The present study evaluated the quality of MCQs used in the Pharmacology semester-end examination conducted in June

2025. The MCQ item analysis were done through difficulty index, discrimination index, distractor effectiveness, and reliability. The findings provide important insights into the overall quality of assessment and areas requiring refinement at the newly established Faculty of Undergraduate Medicine providing Bachelor of Medicine and Bachelor of Surgery (MBBS) degree in the country.

The majority of MCQs 30 (60%) had a difficulty index between 0.26 and 0.75. The difficulty index in these range is considered as a question with the right difficulty[4]. Therefore, these questions may be retained and used in future exams on the same topic subsequently, as per the previous finding[4,5,9]. The questions with a proven right difficulty index may be used for future exams so that the quality of the examination being conducted is good, as claimed by many authors[6,7,10]. However, 12 (24%) MCQs were categorized as very difficult, which may suggest content beyond the expected competency level, ambiguous stems, or inadequate alignment with taught learning objectives, a common reason stated by many previous studies[2,3,5]. All the authors recommend that these questions need either to be revised or discarded for subsequent examination, as this will cause most students to fail[6,7]. Mostafa et al,2026[11] suggested that if revised, again the question would require revalidation to have right difficulty index for possible incorporation in subsequent examinations. Similarly, a smaller proportion of 8 (16%) MCQs were easy. Similar findings were reported in other studies as well[12,13] where a small portion of easy questions needs to be included as a general rule and researchers found it useful for reinforcing core concepts[12,13,14]. But some argued its use should be limited to avoid reducing the discriminatory capacity of the examination[5]. Therefore, some are in favour of revision or discard for inclusion in subsequent examination[4] while others in favour of keeping it[5]. If revised, it needs to be analysed for future inclusion[6].

Item discrimination analysis was performed using the Shrinivas et al. (2023) classification, which defines DI ≥0.50 as excellent, 0.40–0.49 as good, 0.20–0.39 as fair, and ≤0.20 as poor. On this basis, 32 (64%) MCQs had good to excellent discrimination indices. Similar findings were reported by other researchers[5,6,12], indicating that these MCQs were effective in differentiating between high- and low-performing students[1]. These MCQs are considered good quality and may be put in a question bank and may be used for future examination for the same topic, as suggested by Mostafa etal, 2026. But 10 (20%) of the MCQs were found to be fair, which may be used or revised based on the type of examination we conduct, as suggested by Alemu et al 2024[13]. Nevertheless, 8 (16%) of MCQs demonstrated poor discrimination, which may reflect flawed construction, mis keyed answers, or unclear wording. Such items require revision or removal to improve overall test quality, as suggested by most of the researchers[1,2,12].

Distractor effectiveness analysis demonstrated that 67% of items had no non-functioning distractors, indicating generally effective distractor constructions. However, 30 (20%) distractors were non-functioning, while 15(10%) had two non-functioning distractors and 5(3%) had three or more non-functioning distractors. The presence of items with multiple non-functioning distractors indicates that less than 5% of the students selected it, and it is considered non-functional, and this may be replaced by other distractors[5,15]. Effective distractors are essential for enhancing item discrimination and preventing cueing, as claimed by Rana et al. 2024[12].

The reliability analysis of the study yielded a Cronbach's alpha value of 0.78, indicating acceptable internal consistency for summative assessment purposes. The overall reliability of 50 MCQs used in the basic Pharmacology semester end examination was good and acceptable standard and quality. This suggests that the MCQ paper was reasonably consistent in measuring the intended construct[5,12,16]. Regular item analysis, refinement or replacement of poorly performing items can further enhance reliability[11].

Compared with similar item analysis studies from newly established or low-resource medical schools, these findings are broadly consistent. Alemu et al. (2024) and Rana et al. (2024) reported similarly distributed difficulty and discrimination profiles in early-stage curricula, suggesting that an acceptable level of MCQ quality can be achieved even in new institutions, while highlighting the continued need for faculty development in item construction[12,13]. The relatively higher proportion of very difficult items (24%) in the present study may reflect the limited prior experience of faculty in writing MCQs calibrated to the level of second-year students, and warrants targeted faculty development.

Overall, the findings underscore the importance of routine item analysis in strengthening assessment quality. Establishing a validated question bank comprising well-performing MCQs would possibility contribute to maintaining fairness, validity, and reliability in future Pharmacology examinations. Continuous faculty development in MCQ construction and periodic review of assessment tools are recommended to sustain high standards in undergraduate medical education.

LIMITATIONS

The cohort size was small (n = 24), as this is the only MBBS batch currently enrolled at the only medical institution in Bhutan offering undergraduate training; no additional cohort was available to expand the sample. This small cohort limits the statistical power of the discrimination index calculations: with only seven students per extreme group, the DI estimates may be unstable and should be interpreted with caution. The 5% NFD threshold, commonly applied to large cohorts, translates to fewer than 1 or 2 students in this sample (effectively 0 or 1 student), which is an extremely low threshold and may overestimate distractor effectiveness; this is acknowledged as a limitation.

The cross-sectional design and single-centre setting limit generalizability, which is unavoidable given that this is the only institution in Bhutan providing undergraduate medical education. Furthermore, only one of four semester papers (Pharmacology) was analysed; conclusions about overall assessment quality across the programme cannot be drawn from this single paper alone. Future studies should include all four papers and, where possible, combine data across cohorts as the programme matures.

CONCLUSION

This study demonstrates that the majority of MCQs used in the Pharmacology semester-end examination were of acceptable quality, with appropriate difficulty levels, satisfactory discriminatory power, and acceptable internal consistency. However, a proportion of items exhibited poor discrimination and non-functioning distractors, highlighting the need for review, refinement or total replacement. The findings reinforce the importance of routine item analysis as an essential component of quality assurance in summative assessment. Regular evaluation and revision of MCQs will facilitate the development of a robust, validated question bank, thereby enhancing the validity, reliability, and fairness of assessments in undergraduate medical education. Continuous faculty training in MCQ construction and psychometric evaluation is recommended to sustain and further improve assessment standards.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Faculty of Undergraduate Medicine, Khesar Gyalpo University of Medical Sciences of Bhutan, for their support in conducting this study. We extend our appreciation to the Department of Pharmacology for facilitating access to examination data. We also acknowledge the Institutional Review Board (IRB) of KGUMSB for granting ethical waiver. Finally, we thank the MBBS students who participated in the semester-end examination, whose performance data made this analysis possible.

REFERENCES

1. Htoon KZ, Aung YP. Item analysis of multiple-choice questions in summative assessment for professional examination I of an outcome-based integrated MBBS curriculum. Int J Res Med Sci. 2024;12(5):1451–6. [Full Text] [DOI]

2. Farooq M uz Z, Mashood S. Quality Assurance of Multiple-Choice Questions Test Through Item Analysis. Life Sci. 2023;4(4):7. [Full Text] [DOI]

3. Konakci S. Item Analysis in Multiple Choice Questions: A Study on Question Difficulty and Authors' Evaluation. J Basic Clin Heal Sci. 2024;8(2):490–7. [Full Text]

4. Kiyak Ys, Coskun O, Budakoglu Ii, Uluoglu C. Psychometric Analysis of the First Turkish Multiple-Choice Questions Generated Using Automatic Item Generation Method in Medical Education. Tıp Eğitimi Dünyası. 2023;22(68):154–61. [Full Text]

5. Shin S, Choi J, Hong E, Lee M, Lee M. Generating multiple-choice questions using reverse engineering techniques. Med Educ Online. 2026;31(1):1–18. [PubMed] [Full Text] [DOI]

6. Al-Rukban M. Guidelines for the construction of multiple choice questions tests. J Fam Community Med. 2006;13(3):125–33. [PubMed] [Full Text]

7. Srinivas M, Netharakere C, Udaykumar P, Pereira N. Item analysis of multiple-choice questions in pharmacology among medical undergraduates. Natl J Physiol Pharm Pharmacol. 2023;14(5):1. [Full Text] [DOI]

8. Eleragi AMS, Miskeen E, Hussein K, Rezigalla AA, Adam MIE, Al-Faifi JA, et al. Evaluating the multiple-choice questions quality at the College of Medicine, University of Bisha, Saudi Arabia: a three-year experience. BMC Med Educ. 2025;25(1):2–9. [PubMed] [Full Text] [DOI]

9. Obon AM, Rey KAM. Analysis of Multiple-Choice Questions (MCQs): Item and Test Statistics from the 2nd Year Nursing Qualifying Exam in a University in Cavite, Philippines. Abstr Proc Int Sch Conf. 2019;7(1):499–511. [Full Text] [DOI]

10. Fozzard N, Pearson A, Du Toit E, Naug H, Wen W, Peak IR. Analysis of MCQ and distractor use in a large first year Health Faculty Foundation Program: Assessing the effects of changing from five to four options. BMC Med Educ. 2018;18(1):1–10. [PubMed] [Full Text] [DOI]

11. Mustafa S, Hamid OE. Psychometric item/question analysis of multiple-choice questions in fixed prosthodontics exam. BMC Med Educ. 2026;26(1). [PubMed] [Full Text] [DOI]

12. Rana M. Raoof. Evaluation of Anatomy Multiple Choice Questions for First and Second-year Students in the College of Medicine, University of Mosul. Ann Coll Med Mosul. 2024;46(1):128–36. [Full Text] [DOI]

13. Alemu AT, Tesfa H, Mulugeta A, Fenta ET, Belay MA. Quality of multiple choice question items: item analysis. Int J Sci Reports. 2024;10(6):195–9. [Full Text]

14. Adiga MNS, Acharya S, Holla R. Item Analysis of Multiple-Choice Questions in Pharmacology in an Indian Medical School. J Heal Allied Sci NU. 2021;11(03):130–5. [Full Text] [DOI]

15. Gomboo A, Gomboo B, Munkhgerel T, Nyamjav S, Badamdorj O. Item Analysis of Multiple Choice Questions in Medical Licensing Examination. Cent Asian J Med Sci.

2019;5(2):141–8. [Full Text] [DOI]

16. Lamees A, Sajad S. Item analysis of multiple-choice questions in an undergraduate surgery course: An assessment of an assessment tool. Sanamed. 2024;19(2):163–71. [Full Text] [DOI]

AUTHORS CONTRIBUTION

Following authors have made substantial contributions to the manuscript as under:

KT: Principal investigator, concept, design of the protocol, acquisition of data, data analysis/interpretation, drafting/critically reviewing the paper, giving approval for the final version to be published.

GD: Project design, data collection, data analysis, interpretation of results, draft preparation, draft revising, and giving approval for the final version to be published.

KW: Project design, data collection, data analysis, interpretation of results, draft preparation, draft revising, and giving approval for the final version to be published.

Authors agree to be accountable for all respects of the work in ensuring that questions related to the accuracy and integrity of any part of the work are appropriately investigated and resolved.

CONFLICT OF INTEREST

None

GRANT SUPPORT AND FINANCIAL DISCLOSURE

None