

Item Analysis of Physiology Multiple-Choice Questions in an Internal Assessment Among Undergraduate Medical Students in Central India: A Single-Test Post-Hoc Study

Mohammad Ghodke^{1*}, Mohammed Salahuddin², Mohammed Yaser Askari², Syed Badar Daimi³, Azhar Siddiqui⁴

¹Assistant Professor, Department of Community Medicine, JIIU's IIMSR, Jalna, Maharashtra, India

²Associate Professor, Department of Physiology, JIIU's IIMSR, Jalna, Maharashtra, India

³Professor, Department of Physiology, JIIU's IIMSR, Jalna, Maharashtra, India

⁴Professor & Dean, Department of Anatomy, JIIU's IIMSR, Jalna, Maharashtra, India

*Corresponding Author:

Mohammad Ghodke

E-MAIL: dr.msghodke@gmail.com



COPYRIGHT: ©2026 Ghodke et al. This is an open-access journal, and articles are distributed under the terms of the Creative Commons Attribution License CC-BY 4.0. (<https://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.

Date of Submission: 10-04-2026

Date of Review: 14-04-2026

Date of Acceptance: 22-04-2026

ABSTRACT

Introduction: Multiple-Choice Questions (MCQs) are widely used in undergraduate medical assessment as they permit broad content coverage, objectivity in scoring, and evaluation of large groups. Post-test item analysis guides retention or revision of items for future use. **Objectives:** To analyze Physiology MCQs from MBBS Phase I internal assessment using item analysis parameters to generate evidence-based recommendations for revision or retention of items. **Methods:** This study was a single-test post-hoc item analysis of an existing Physiology MBBS Phase I internal assessment conducted at JIIU's Indian Institute of Medical Science and Research Medical College, Maharashtra, India. All students who appeared for the test were included (n=123). The paper comprised 20 single best-answer MCQs, each with one key and three distractors. Item-level analysis was performed on anonymized responses using pre-specified criteria. The KR-20 reliability coefficient was calculated. **Results:** Final analysis included 123 students and 20 MCQs. The mean score was 9.93 ± 2.48 , with a range of 4-16. The mean difficulty index was 0.496, mean discrimination index was 0.303, and mean distractor efficiency was 78.3%, and KR-20 reliability coefficient was 0.35. 4 items were difficult, 11 were of moderate difficulty, and 5 were easy. Good discrimination was observed in 5 items, and 11 items had 100 % distractor efficiency. Using the operational retention rule, 4 items were retained, 16 required revision, and only 2 items met all three desired criteria.

Conclusion: The Physiology MCQs demonstrated acceptable mean difficulty, average discrimination, and good distractor efficiency, however most items required revision before inclusion in departmental MCQ bank.

KEYWORDS: Difficulty index, Discrimination index, Distractor efficiency, Internal assessment, Item analysis, Physiology

INTRODUCTION

Multiple-choice questions (MCQs) are widely used in departmental examinations as they have the advantage of sampling broad domains of knowledge efficiently and hence reliably^[1].

The MCQ format allows teachers to efficiently assess large numbers of candidates and to test a wide range of content^[2].

Well-constructed MCQs are preferred for their objectivity in assessment, comparability in different settings, wide coverage of subject, and minimization of assessor's bias^[3].

Item analysis is a post-examination method for ensuring the validity and reliability of MCQs. It provides feedback about the constructed items and the coverage of the content materials from which items were created^[4].

Failure to adhere to standard item-writing guidelines may render examination questions easier or more difficult than intended^[5].

Common item analysis parameters include the difficulty index (DIF I), which reflects the percentage of correct answers to total responses; the discrimination index (DI), which identifies discrimination between students with different levels of achievement; and distractor efficiency (DE), which indicates whether the distractors in the item are well-chosen or have failed to distract students from selecting the correct answer^[6].

The present study was conducted to analyse the Physiology MCQs in MBBS Phase I internal assessment examination using these item analysis parameters to generate evidence-based recommendations for retention, and revision of items and to strengthen departmental MCQ bank.

METHODS

This study was a single-test post-hoc item analysis of existing Physiology MBBS Phase I internal assessment conducted at JIU's Indian Institute of Medical Science and Research Centre Medical College in Maharashtra, India. The study was conducted over a period of six weeks from 15 Feb 2026 to 31 March 2026. All students who appeared for the test were included (n=123). The primary unit of analysis was an MCQ item (n=20). Student responses were used to calculate item level psychometric indices from anonymized examination data. Incomplete and non-finalized records from anonymized dataset were excluded from analysis. The written question paper included 20 single best-answer MCQs. Each item consisted of one correct response i.e. key and three distractors. Analysis was done at item level and responses were anonymized. Each correct answer response carried 1 mark. There was no negative marking. Responses were coded 1 for correct and 0 for incorrect and entered into Microsoft Excel 2013 (Microsoft Corp., USA) and analysed at item level. Because the objective of the present study was a descriptive psychometric analysis of a single examination paper, inferential statistics were not planned. Descriptive statistics were expressed as mean, standard deviation, range, frequencies, and percentages. Mean difficulty index, mean discrimination index, and mean distractor efficiency were reported to summarize the overall profile and were interpreted together with category-wise distributions and item-level results rather than as stand-alone indicators. The Kuder Richardson 20 (KR-20) reliability coefficient was computed from the finalized binary response matrix to describe the internal consistency of 20-item paper.

The Institutional Ethics Committee (IEC) approval was taken. Difficulty index was defined as proportion of students answering an item correctly. Difficulty was

categorized as difficult (< 0.30), moderate (0.30 – 0.70) and easy (> 0.70). After ranking students by total score, the upper and lower 27% groups were selected according to Kelley's criterion for extreme-group item analysis^[13]. As 27% of 123 students equals 33.21, the group size was rounded to 33 students in each group. The discrimination index (DI) was calculated as $DI = (H-L)/n$ ^[13], where H is the number of students answering the item correctly in high-achiever group, L is the number answering correctly in low-achiever group, and n is the number of students in each extreme group. The DI ranged from - 1 to + 1. Items were categorized as poor (< 0.20), acceptable (0.20 – 0.39), and good (≥ 0.40) discriminators^[11].

Distractor efficiency is a measure of functionality of each of three distractors per item. A distractor selected by fewer than 5% students is considered as non-functional distractor. Accordingly, items with 0, 1, 2, and 3 non-functional distractors had distractor efficiency values of 100%, 66.7%, 33.3%, and 0 %, respectively^[8].

RESULTS

Final analysis included 123 students and 20 single-best answer MCQs. The upper and lower groups comprised 33 students each. The minimum score was 4 and the maximum score was 16. The mean total score was 9.93 with standard deviation of 2.48. The mean difficulty index was 0.496 (range: 0.122-0.894), the mean discrimination index was 0.303, and the mean distractor efficiency was 78.3% and the KR-20 reliability coefficient for the paper was 0.35 [Table. 1].

Based on predefined classification on item-wise analysis, 4 items (20%) were categorized as difficult, 11 items (55.0%) as moderate, and 5 items (25.0%) as easy [Table. 2]. The most difficult item was Q2 (difficulty index 0.122) and the easiest item was Q4 (difficulty index 0.894) [Table. 1].

With respect to discrimination index (DI), 5 items (25%) demonstrated good discrimination, 10 items (50.0%) showed acceptable discrimination, and 5 items (25%) showed poor discrimination. No item demonstrated negative discrimination [Table. 3]. The highest discrimination index was observed for Q5 and Q12 (0.545 each), followed by Q18 (0.515), whereas item number Q10 had the lowest discrimination index (0.00) [Table. 1].

The mean distractor efficiency was 78.3%. Out of 20 items, 11 items (55.0%) had distractor efficiency of 100%, 5 items (25.0%) had 66.7%, and 4 items (20.0%) had 33.3%. No item had 0.0% distractor efficiency [Table. 4].

Based on pre-specified operational decision rule of moderate difficulty (0.30 – 0.70), discrimination index (DI ≥ 0.20), and distractor efficiency of 100 %, 4 items (Q1, Q5, Q12, and Q19) were considered suitable for retention,

whereas remaining 16 items required revision. No item fulfilled criterion for discard, as none showed negative discrimination [Table. 1].

Overall, only 2 items (10.0%) met all three desired criteria simultaneously, namely moderate difficulty (0.30- 0.70), good discrimination (DI \geq 0.40) and 100 % distractor efficiency. The 18 items (90.0 %) did not meet all three criteria [Table. 5].

Item No.	Difficulty Index (p value)	Discrimination index (DI)	Distractor efficiency (DE)	Decision
Q1	0.415	0.333	100.00	Retain
Q2	0.122	0.061	100.00	Revise
Q3	0.390	0.424	66.67	Revise
Q4	0.894	0.242	33.33	Revise
Q5	0.439	0.545	100.00	Retain
Q6	0.293	0.394	100.00	Revise
Q7	0.163	0.242	100.00	Revise
Q8	0.675	0.061	100.00	Revise
Q9	0.268	0.424	100.00	Revise
Q10	0.350	0.000	100.00	Revise
Q11	0.764	0.182	33.33	Revise
Q12	0.423	0.545	100.00	Retain
Q13	0.756	0.242	66.67	Revise
Q14	0.740	0.333	33.33	Revise
Q15	0.610	0.242	66.67	Revise
Q16	0.358	0.182	100.00	Revise
Q17	0.577	0.364	66.67	Revise
Q18	0.496	0.515	33.33	Revise
Q19	0.472	0.364	100.00	Retain
Q20	0.724	0.364	66.67	Revise

Table 1: Item-wise analysis summary

Difficulty category	Classification criteria	Item numbers	n	Percentage
Difficult	< 0.30	Q2, Q6, Q7, Q9	4	20.0
Moderate	0.30 - 0.70	Q1, Q3, Q5, Q8, Q10, Q12, Q15, Q16, Q17, Q18, Q19	11	55.0
Easy	>0.70	Q4, Q11, Q13, Q14, Q20	5	25.0
Total	-	20 items	20	100.0

Table 2: Distribution of difficulty index (p value) categories (n=20)

Discrimination category	Number of items	Percentage
Good (\geq 0.40)	5	25.0
Acceptable (0.20 - 0.39)	10	50.0
Poor (< 0.20)	5	25.0
Negative (-)	0	0.0
Total	20	100

Table 3: Distribution of discrimination index (DI) categories (n=20)

Distractor efficiency (%)	Number of items	Percentage
100	11	55.0
66.7	5	25.0
33.3	4	20.0
0.0	0	0.0
Total	20	100

Table 4: Distribution of distractor efficiency (DE) categories (n=20)

Criterion status	Number of items	Percentage
Meet all three desired criteria	2	10.0
Did not meet all three desired criteria	18	90.0
Total	20	100.0

Table 5: Items meeting all three desired criteria simultaneously (n=20)

Desired criteria: Moderate difficulty index (0.30 - 0.70), Good discrimination index (\geq 0.40) and distractor efficiency (100%)

DISCUSSION

This analysis showed that physiology paper had a reasonable average difficulty level and good distractor functioning, but item quality was uneven at individual-question level. Only four items satisfied the operational retention rule, and only two items simultaneously achieved moderate difficulty, good discrimination, and full distractor efficiency. Therefore, the overall paper should not be interpreted as uniformly strong. The findings support selective retention of a small subset of items and substantial revision of the remaining MCQs before reuse in the departmental question bank.

Gajjar *et al.*^[3] reported a lower mean discrimination index (0.14), but higher mean distractor efficiency (88.6%).

Rezigalla *et al.*^[4] reported 69.5% items with acceptable difficulty but a higher mean discrimination index (0.46).

Kheyami *et al.*^[6] reported mean difficulty indices ranging

from 36.70% to 73.14%, mean discrimination indices from 0.20 to 0.34, and mean distractor efficiency from 66.5% to 90%. Mahjabeen *et al.*^[9] found 81% items in the acceptable difficulty range with mean discrimination index of 0.35 ± 0.16 and mean distractor efficiency of 63.55 ± 27.47 . Kumar *et al.*^[10] reported 82% items with good or acceptable difficulty, a mean difficulty index of 55.32 ± 7.4 , and mean discrimination index of 0.31 ± 0.12 . Patil *et al.*^[11] reported a mean difficulty index 38.3%, a mean discrimination index 0.27, and a mean distractor efficiency 82.8%. Bhat and Prasad^[12] reported mean difficulty index 53.22, mean discrimination index 0.26, and mean distractor efficiency of 78.32%. Q10 had a discrimination index of 0.00, indicating that the proportions of correct responses in upper and lower groups were equal; therefore, the item did not differentiate between high and low performing students. Although Q10 had moderate difficulty and 100 % distractor efficiency, its lack of discriminatory power suggests that it should be reviewed for stem clarity, alignment with taught content, and key accuracy. Similarly, Q2 and Q8, which showed very low discrimination, need focused content and construction review before future use. The present study findings fall within the broad range reported in the literature for item analysis indices. It also supports structured revision of a large proportion of items before their inclusion in the departmental question bank.

CONCLUSION

The analyzed Physiology paper showed acceptable mean difficulty, average discrimination, and reasonably good distractor efficiency, but only a small number of items were suitable for retention and the internal consistency of 20-item paper was low. These findings support systematic post-examination item analysis followed by targeted revision, rather than direct reuse of most items, to strengthen future departmental assessments and build a departmental MCQ bank.

Limitations:

This study was limited to single internal assessment paper from one institution, and one subject area, and it included a small number of items i.e. (n=20). The analysis was restricted to single examination and did not include repeat testing, test-retest evidence, or longitudinal item performance across batches. The manuscript also did not include formal blueprinting or content validity assessment, qualitative review of item-writing flaws, or cognitive-level classification according to Bloom's taxonomy. In addition, other psychometric indicators such as point-biserial correlation were not computed. Accordingly, the findings should be interpreted as an item-level psychometric analysis of one examination

rather than a comprehensive validation of the broader assessment program.

DISCLOSURE

Conflict of interest: None declared.

Funding: Nil.

References

1. Sim SM, Rasiah RI. Relationship Between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. *Annals of the Academy of Medicine, Singapore*. 2006; 35 (2) :67-71 . Available from: <https://doi.org/10.47102/annals-acadmedsg.v35n2p67>
2. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*. 2009; 9 (1) . Available from: <https://doi.org/10.1186/1472-6920-9-40>
3. 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 Journal of Community Medicine*. 2014; 39 (1) :17-20 . Available from: <https://doi.org/10.4103/0970-0218.126347>
4. Rezigalla AA, Eleragi AMESA, Elhussein AB, Alfaifi J, Alghamdi MA, Al Ameer AY, *et al.* Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*. 2024; 24 (1) :445 . Available from: <https://doi.org/10.1186/s12909-024-05433-y>
5. Rush BR, Rankin DC, White BJ. The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education*. 2016; 16 (1) :250 . Available from: <https://doi.org/10.1186/s12909-016-0773-3>
6. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal*. 2018; 18 (1) :e68-e74 . Available from: <https://doi.org/10.18295/squmj.2018.18.01.011>
7. Mustafa AEM. Evaluation of the item analysis of multiple-choice pediatric exams: a college of medicine departmental review. *Bahrain Medical Bulletin*. 2024; 46 (3) :2363-2369 . Available from: <https://doi.org/10.21203/rs.3.rs-1994965/v1>
8. Namdeo SK, Rout SD. Assessment of functional and non-functional distracter in an item analysis. *International Journal of Contemporary Medical Research*. 2016;3(7):1891-1893.
9. Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S, *et al.* Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of Pakistan Institute of Medical Sciences*. 2017; 13 (4) :310-315 . Available from: <https://doi.org/10.48036/apims.v13i4.9>
10. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*. 2021; 77 (Suppl 1) :S85-S89 . Available from: <https://doi.org/10.1016/j.mjafi.2020.11.007>

11. Patil R, Palve SB, Vell K, Boratne AV. Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. *International Journal of Community Medicine and Public Health*. 2016; 3 (6) :1612-1616 . Available from: <https://doi.org/10.18203/2394-6040.ijcmph20161638>
12. Bhat SK, Prasad KHL. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology. *Indian Journal of Ophthalmology*. 2021; 69 (2) :343-346 . Available from: https://doi.org/10.4103/ijo.ijo_1610_20
13. Kelley TL. The selection of upper and lower groups for the validation of test items.. *Journal of Educational Psychology*.

1939; 30 (1) :17-24 . Available from:
<https://doi.org/10.1037/h0057123>

How to cite this article: Ghodke M, Salahuddin M, Askari MY, Daimi SB, Siddiqui A. Item Analysis of Physiology Multiple-Choice Questions in an Internal Assessment Among Undergraduate Medical Students in Central India: A Single-Test Post-Hoc Study. *Perspectives in Medical Research* 2026; 14(1):79-83 DOI: [10.47799/pimr.1401.26.46](https://doi.org/10.47799/pimr.1401.26.46)