ISSN: 2581 – 3633

International Journal of Medical Science and Advanced Clinical Research (IJMACR)

Available Online at: https://www.ijmacr.com

Volume – 9, Issue – 6, November – 2026, Page No.: 316 – 321

Item Analysis of Multiple-Choice Questions-A Retrospective Tool for a Prospective Database

[1]Dr. Gayathri Chelamkuri, M.D., Associate Professor, Department of Biochemistry, Kurnool Medical College, Kurnool-518002, Andhra Pradesh

[2]Dr. N. Lakshmanna, M.D., Assistant Professor, Department of Biochemistry, Kurnool Medical College, Kurnool-518002, Andhra Pradesh

[3]Dr. A Padma Vijaya Sree, M.D., Professor, Department of Biochemistry, Viswabharathi Medical College, Penchekalpadu, Kurnool, 518467, Andhra Pradesh

Corresponding Author: Dr. Gayathri Chelamkuri, M.D., Associate Professor, Department of Biochemistry, Kurnool Medical College, Kurnool-518002, Andhra Pradesh

How to citation this article: Dr. Gayathri Chelamkuri, Dr. N. Lakshmanna, Dr. A Padma Vijaya Sree, "Item Analysis of Multiple-Choice Questions-A Retrospective Tool for a Prospective Database", IJMACR– November – 2026, Volume – 9, Issue – 6, P. No. 316 – 321.

Open Access Article: © 2026, Dr. Gayathri Chelamkuri, et al. This is an open access journal and article distributed under the terms of the creative common's attribution license (http://creativecommons.org/licenses/by/4.0). Which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

Type of Publication: Original Research Article

Conflicts of Interest: Nil

Abstract

Introduction: Assessment is determining whether the stated educational objectives have been achieved and is integral to educational experience. The introduction of Competency Based Medical Education (CBME) in 2019 emphasized the development of required competencies in an organized manner through learner centered approach. Appropriate assessment tools are needed to assess the competence pertaining to all the learning domains. Graduate Medical Education Regulations (GMER) has accorded MCQs a weightage of not more than 20% of the total theory marks in summative assessment. MCQs provide an objective evaluation and can also assess higher cognitive level. Validity and Reliability are of vital importance for any assessment tool. The post-test item analysis of MCQs can provide input on validity and reliability of items that can be prospectively used in designing the question bank.

Aims and Objectives: The present study is intended to perform the post-test validation of MCQs in the Biochemistry Internal assessments for Phase-I MBBS students.

Method: 80 MCQs from Biochemistry internal assessments from 2023-2024 batch Phase I MBBS were analyzed for Difficulty Index (DI), Discrimination Index and Distractor Effectiveness, and the statistical analysis is performed using Microsoft Excel.

Results: Analysis of 80 MCQs showed, 58 (72.5%) test items had good/acceptable Difficulty Index, 2 (2.5 %) were too difficult and 20 (25%) were too easy. The

Discrimination Index score was poor for 13 (16.25%), acceptable for 2 (2.5%), good for 25 (31.25%) and excellent for 40 (50%) respectively. Of the 240 distractors for 80 MCQs, 164 (68.3%) were functional and 76 (31.7%) were non-functional.

Conclusion: Item analysis provides valuable data on improving the quality of MCQs, building a reliable question bank, minimizing the non-functional distractors and discarding any defective test items that compromise the objective of assessment.

Keywords: Assessment, MCQs, Difficulty Index, Discrimination Index, Distractor Effectiveness

Introduction

- Multiple Choice Questions (MCQs) have exerted their widespread dominance in gateway entrance examinations and licensing exams over the years. The scope of MCQs have drastically expanded for use in both continuous low-stake formative assessments and high-stakes end of phase examinations. Graduate Medical Education Regulations (GMER) has accorded Multiple-choice questions (MCQs) a weightage of not more than 20% of the total theory marks in the CBME curriculum.
- MCQs offer a quick, objective evaluation and can be used at the end of instructional practice for pre-lecture and post-lecture readiness among the students to drive flipped classrooms as well.
- Use of multiple-choice item serves as a quick tool for formative evaluation, determine areas which require emphasis, reinforcement or alteration in teaching methodology and perhaps learning aids.
- Though there were widespread concerns that MCQs can only test recall of factual knowledge ('knows' of Miller's Pyramid) likely to encourage surface learning, Higher-order thinking (H.O.T) skills can also be tested by appropriate questions.
- In Competency-Based Medical Education (CBME), Multiple-Choice Questions (MCQs) are transitioning from simple rote-memorization tests to sophisticated tools capable of measuring clinical reasoning. When aligned with the principles of CBME—such as those instituted by the National Medical Commission (NMC) MCQs have a distinct scope, utility, and set of trade-offs. The use of Scenario-based MCQs or Case-based MCQs address clinical application and analysis moving to 'knows how' of Miller's Pyramid.
- Five criteria that are used to evaluate the assessment tools include: Validity, Reliability, Educational impact, Acceptability and Cost effectiveness.

Validity of the MCQs can be ensured by:

Pre-validation	Post-validation
Qualitative analysis	Quantitative analysis / Item analysis test
Proof-reading the test before it is administered	The process of analysing an item after it has appeared in an examination
Relevance with respect to the learning outcomes	Difficulty index or Facility value or *p*-value
Grammar of construction	Discrimination index or *d*-value
Clarity, brevity and appropriateness of stem	Distractor effectiveness, DE
Plausibility of all distractors	

©2026, IJMACR

An estimate as to the level of cognition tested by them	

Item Difficulty index / Facility Value or p-Value [2,4]

- It is the relative frequency with which examinees choose the correct response.
- The *percentage of students who answers the item correctly (H + L groups)*

$$P = \frac{H + L}{N} \times 100$$

30 – 70%	Acceptable
50 – 60%	Ideal
> 70 %	Very Easy
< 30%	Very Difficult

Discrimination index, *d*-value[2,4]

It is the ability of the test item to differentiate between good and not so good students.

$$d = \frac{H - L}{N} \times 2$$

≤ 0.20	Poor
0.21 – 0.24	Acceptable
0.25 – 0.35	Good
≥ 0.36	Excellent

Where *H* and *L* represent No. of students answering the item correctly in the high achieving group and low achieving group respectively

N: Total No. of students in the two groups (including the non-responders)

Distractor effectiveness or functionality [2.4]

- Distractor analysis is aimed to identify functioning and non-functioning distractor
- Any distractor in the item which has not attracted even 5% of the total response is said to be non-functional distractor.
- MCQs require good number of plausible distractors to achieve reliability.
- Based on number of NFDs in an item, DE ranges from 0 to 100%.

No of NFDs in a test item	DE
0	100%
1	66.6%
2	33.3%
3	0

Aims & Objectives

The present study is aimed at the post-test validation of MCQs from the internal assessments conducted in the department of Biochemistry.

Objectives

1. To calculate the parameters: Difficulty index (*p*-value), Discrimination index (*d*-value) and Distractor Effectiveness (DE) of test items
2. To interpret the values as per the standards

Materials & Methods

The summative examination conducted by the university comprised 20 single-best-response MCQs with no provision for negative marking. In the current study, 80 pre-validated MCQs from all the four Internal assessments conducted to 250 Phase-I MBBS (2023-2024) students in the department of Biochemistry were included. The number of MCQs included in each internal assessment is 20. MCQs consisted of a stem with single best response and three distractors. One mark awarded for correct response and zero for wrong response, no negative marking.

All the scores of the students for each assessment were noted in Microsoft Excel and data is sorted to rank students in order of merit. The top third students are taken as High achievers ***(H)*** and the bottom third as Low

©2026, IJMACR

achievers ***(L)***. Item analysis is done calculating the Difficulty Index, Discrimination Index and Distractor functionality. Statistical analysis is performed by computing the Mean and Standard Deviation.

Table 1: Descriptive analysis of the study data

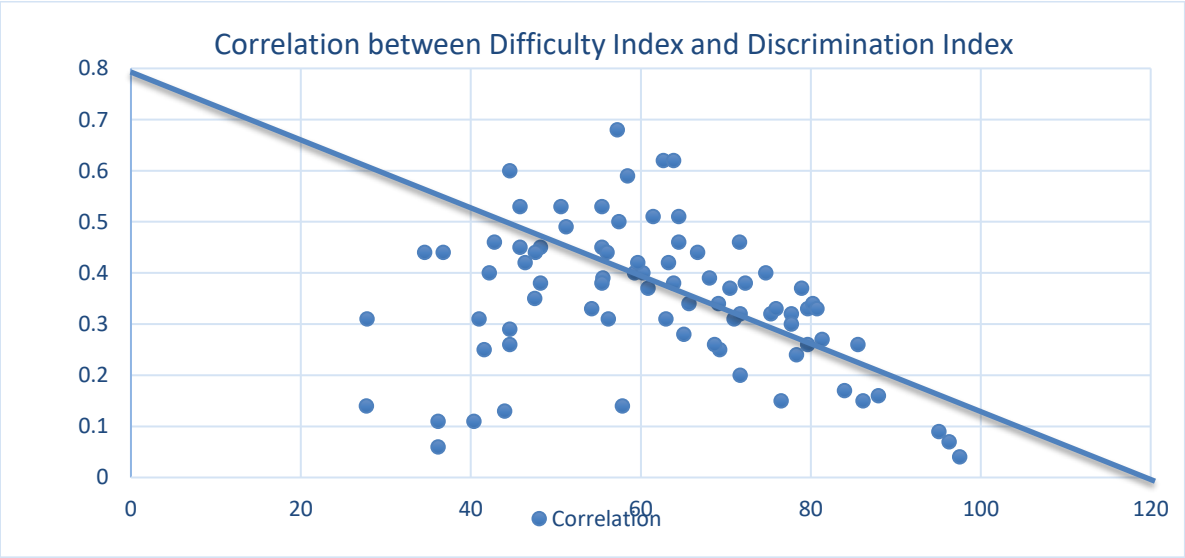
No of test items	80
No of students	250
Overall mean test score $\pm S.D$	12.25 ± 14.14
Median	12
Range of test scores	3 - 20
Mean test score $\pm S.D$ (High achievers)	15.58 ± 2.97
Mean test score $\pm S.D$ (Low achievers)	8.74 ± 0.89

Table 2: Distribution of Test items as per the Difficulty Index or p value

P-value	Interpretation	No of test items (%)	Mean ± S. D
< 30	Too difficult	2 (2.5%)	27.74 ± 0.02
30 - 70	Good / Acceptable	58 (72.5%)	55.65 ± 9.53
> 70	Too easy	20 (25%)	82.45 ± 10.66

Table 3: Distribution of Test items as per the Discrimination Index or d value

d-value	Interpretation	No of test items (%)	Mean ± S. D
≤ 0.20	Poor	13 (16.5%)	0.12±0.06
0.21 - 0.24	Acceptable	2 (2.5%)	0.24±0
0.25 – 0.34	Good	25 (31.25%)	0.30±0.04
≥ 0.35	Excellent	40 (50%)	0.45±0.07



Correlation coefficient, r = (-0.266)

©2026, IJMACR

Table 4: Distribution of Test items as per the Distractor Effectiveness (DE)

Total number of Distractors for 80 MCQs: 240			
Functional Distractors	Non-functional distractors		
164 (68.3%)	76 (31.7%)		
No of test items….			
without NFD (DE: 100%)	with 1 NFD (DE: 66.6%)	with 2 NFD (DE: 33.3%)	with 3 NFD (DE: 0)
30 (37.5%)	28 (35%)	18 (22.5%)	4 (5%)

Discussion

Item Analysis is a relatively simple method for post-validation of the MCQs. Table No.1 shows descriptive analysis of the study data. An ideal single best response item having four options will be the one with an acceptable difficulty index (30% to 70%), good discrimination index (> 0.25) and maximum distractor effectiveness (100%). In the present study (table no.2), a total of 58 (72.5%) items had acceptable level of p-value and a 20 (25%) are too easy with respect to p-value. These easy questions included in the beginning of question paper in summative assessment may serve as warm-up questions and can be used often in formative assessment. Their educational value therefore depends on their intended purpose within assessment.

Item discrimination is a measure of the ability of the item to discriminate between good students and not so good students. From table No 3, 40 (50%) test items had an excellent d-value and 25 (31.25%) had a good d-value. As a tool for evaluation of classroom performance, MCQs with good d-value might be sufficient as the main aim is to ensure that all students have learnt the lesson correctly. A good to excellent d-value might be of useful for tests summative evaluation and assessments for ranking students.

An important aspect of framing an MCQ is designing plausible and homogenous distractors. From table No 4, the present study 30 test items had distractor effectiveness (DE) 100%, 28 had 66.6%, 18 had 33.3% and 4 had 3 non-functional distractors with DE of 0. Presence of non-functional distractors makes the test item easy and leads to a greater possibility of guessing the correct response. Plausible and homogenous distractors are essential for preventing unintended clues and reducing the likelihood of guessing the correct response.

Importantly, no single item-analysis parameter should be interpreted in isolation. An item may be easy but still useful or may have acceptable difficulty but poor discrimination. Similarly, a question with non-functional distractors may require revision rather than immediate rejection. Therefore, item analysis should be used as decision-support tool for reviewing, revising and classifying MCQs rather than as an automatic mechanism for accepting or discarding items.

An r of -0.266 represents a negative direction and weak correlation between p-value and d-value. As the item becomes easier, the discrimination index tends to decrease slightly. Within the dataset, easier questions tended to have somewhat lower discrimination whereas questions with lower difficulty indices tended to show somewhat higher discrimination.

©2026, IJMACR

Conclusion

Item analysis provides a simple, objective and educationally useful method for post-validation of MCQs. Item analysis helps to create an item bank of tested MCQs with known levels of difficulty and discriminating power, free of constructional errors and having functional distractors. In the present study, most items demonstrated an acceptable level of difficulty, while a substantial proportion showed good to excellent discrimination. However, the presence of non-functional distractors and poorly discriminating items identified specific areas requiring review and improvement. Their combined interpretation can guide faculty in identifying items that may be retained, revised or reconsidered for future assessments.

Therefore, item analysis transforms a retrospective examination dataset into a prospective quality-improvement tool.

Limitations of the Study

1. Reliability, a measure of internal consistency is not studied.
2. Interrelationships among the *p*-value, *d*-value and DE is not evaluated.
3. Response frequency is not studied.

References

1. Bhattacherjee Sharmistha *et al* Evaluation of Multiple-choice Questions by Item Analysis, from an Online Internal Assessment of 6th semester Medical Students in a Rural Medical College, West Bengal Indian Journal of Community Medicine 47(1): p 92-95, Jan-Mar 2022
2. Dharmedra Kumar *et al* Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Med J Armed Forces India, 2021 Feb; 77 S85-S89
3. Rupal M Patel Use of Item analysis to improve quality of Multiple Choice Questions in II MBBS Journal of Education Technology in Health Sciences, January-April, 2017; 4(1):22-29
4. N. Ananthakrishnan: The value of item analysis in classroom teaching, Indian Journal of Medical Education

©2026, IJMACR