International Journal of Research in Medical Sciences
Aung YP et al. Int J Res Med Sci. 2025 Apr;13(4):1383-1390
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20250957

Original Research Article

Item analysis of type "A" multiple choice questions of biochemistry in GIM module, year 1 MBBS program

Ye Phyo Aung[1]*, Yan Naing Soe[2], Mo Mo Than[2]

[1]Department of Medical Education, Defence Services Medical Academy, Yangon, Myanmar
[2]Department of Biochemistry, Defence Services Medical Academy, Yangon, Myanmar

Received: 30 January 2025
Revised: 04 March 2025
Accepted: 07 March 2025

***Correspondence:**
Dr. Ye Phyo Aung,
E-mail: yephyoaung@dsma.edu.mm

ABSTRACT

Background: Medical education is continually evolving to meet the demands of healthcare and scientific advancements. The Defence Services Medical Academy (DSMA) in Myanmar implemented an outcome-based curriculum in 2017. The genetics, immunology, and molecular medicine (GIM) module is a critical component of the first-year MBBS program, laying the foundation for knowledge in biochemistry and related sciences.
Methods: This study analyzed the quality of type "A" multiple choice questions (MCQs)- 28 items from the Biochemistry section of the GIM module's end-module assessment. The evaluation focused on four key indices: difficulty index (P), discrimination index (D), distractor efficiency (DE), and Kuder-Richardson formula 20 (KR-20) for reliability.
Results: Of the 28 items, 21 (75%) were classified as having average difficulty, while 5 (18%) were deemed too difficult. Regarding discrimination, 12 (43%) items displayed very good discrimination, while 8 (29%) had poor discrimination, indicating the need for revision. Nearly 70% of the MCQs had fully functional distractors. Overall, the Biochemistry questions showed moderate reliability (KR-20 score =0.682), with three items (11%) recommended for rejection due to poor performance.
Conclusions: These findings highlight the necessity of regular item analysis and revision to ensure the quality and fairness of assessments. Faculty development and active learning strategies are essential to improving the overall reliability and effectiveness of MCQs in medical education.

Keywords: Difficulty index, Discrimination index, Distractor efficiency, Item analysis, KR20, One best answer, Type A MCQ

INTRODUCTION

Medical education is continuously evolving to align with the changing demands of healthcare systems and scientific advancements. This transformation reflects the growing recognition of the need for integrated knowledge, skills, and attitudes in medical practice. In response to this global trend, the Defence Services Medical Academy (DSMA) in Myanmar implemented a new curriculum in 2017, transitioning from a traditional approach to an outcome-based, integrated medical education program. This shift aligns with the World Federation for Medical Education (WFME) basic medical education standards, emphasizing the importance of incorporating knowledge, skills, and professional behaviors to enable medical students to deliver competent care.[1]

The six-year MBBS program at DSMA includes a series of well-structured modules in the first year, one of which is the Genetics, Immunology, and Molecular medicine (GIM) module. The GIM module plays a critical role in establishing a foundation in the basic sciences, including

Biochemistry, Genetics, Molecular biology, Anatomy, Microbiology, Pathology, and Pharmacology- subjects essential for contemporary clinical practice. Integrating these disciplines within the GIM module ensures that students develop a comprehensive understanding of genetic and immunological concepts, enabling them to apply this knowledge in clinical contexts. The study of molecular genetics, in particular, emphasizes key topics such as DNA structure, replication, and the central dogma of molecular biology, which are crucial for understanding the molecular basis of genetic disorders.

A key component of medical education is the assessment of students' knowledge and understanding of the content covered in various modules. The GIM module employs multiple teaching and learning strategies, including lectures, practical sessions, tutorials, and problem-based learning (PBL), supported by directed self-learning (DSL) and self-directed learning (SDL). These methods enhance students' engagement with the material and foster deep learning. Assessments in the GIM module, including multiple-choice questions (MCQs), are designed to align with the intended learning outcomes. MCQs, particularly the single-best answer or type "A" MCQs, are widely used to evaluate student performance due to their ability to assess both factual knowledge and higher-order cognitive skills, such as interpretation, synthesis, and application.[2-4]

MCQs are a significant assessment tool in medical education because they provide an objective, reliable, and efficient means of evaluating large groups of students. Constructing high-quality MCQs requires careful consideration and expertise to ensure they assess not only recall but also critical thinking and application. A well-constructed MCQ consists of a stem (the question), a correct answer (the key), and several distractors (incorrect options). The distractors play a crucial role in ensuring that only students who thoroughly understand the content can select the correct answer, while those with less confidence are drawn to the incorrect options. The quality of MCQs is assessed through item analysis, which evaluates the reliability and validity of each question.[5,6]

Item analysis is a post-examination process that provides valuable insights into the effectiveness of test items. It measures three key indices: the difficulty index (P), the discrimination index (D), and distractor efficiency (DE). Each of these indices provides distinct information on how well an MCQ functions in assessing student knowledge and differentiating between high and low performers.[7-10]

The difficulty index (P) refers to the percentage of students who correctly answered a particular item, ranging from 0% (if no student answered correctly) to 100% (if all students answered correctly). An ideal difficulty index falls between 30% and 70%, indicating that the item is neither too easy nor too difficult. Items with a difficulty index below 30% are considered too difficult, whereas those above 70% are deemed too easy. Proper difficulty balance ensures that the exam differentiates between students with varying levels of understanding. However, the inclusion of more difficult items may be appropriate when the goal is to identify top performers.[7-9]

The discrimination index (D), also called the point-biserial correlation (PBS), measures how well an item differentiates between high- and low-achieving students. It ranges from -1.00 to +1.00. A positive discrimination index (0.00 to +1.00) indicates that high-performing students are more likely to answer correctly, while a negative discrimination index (-1.00 to 0.00) suggests that low-performing students answer correctly more often, signaling the need for item revision.[7-10]

Distractor efficiency (DE) assesses how well the distractors function in misleading students who do not fully understand the content. Non-functioning distractors (NFDs) are those selected by fewer than 5% of students, indicating they are too implausible to serve their purpose. Ideally, functional distractors should be chosen by at least 5% of students to effectively challenge those uncertain about the correct answer.[7-9]

The construction and analysis of MCQs are crucial for ensuring the validity and reliability of assessments in medical education. The reliability of an MCQ test can be measured using the Kuder-Richardson Formula 20 (KR-20), a statistical measure used to assess the internal consistency of assessments with dichotomous choices (i.e., items with two possible responses, such as true/false or yes/no). Internal consistency refers to the extent to which all items in a test measure the same construct or trait. A KR-20 score of 0.8 is generally considered the minimum acceptable value; scores below 0.8 may indicate that the exam lacks reliability. However, a high KR-20 score does not necessarily indicate a homogeneous test, as this remains an assumption rather than a definitive conclusion.[11,12]

Together, these indices provide a comprehensive evaluation of assessment quality, ensuring that assessments are both fair and meaningful. MCQs in medical education offer several advantages, including efficient administration, objective scoring, and the ability to assess a broad range of content. However, constructing high-quality MCQs requires significant expertise and meticulous attention to detail. Item analysis serves as a valuable tool for evaluating MCQ quality, helping educators identify problematic items and refine their assessment strategies. By continuously analyzing and improving MCQs, educators can ensure that assessments are fair, reliable, and effective in measuring student learning.[3,8,13]

The GIM module in the year 1 MBBS program at DSMA is designed to provide students with a foundational understanding of genetics, immunology, and molecular biology. Through a well-structured curriculum, diverse teaching and learning activities, and robust assessment methods, the module equips students with the knowledge and skills necessary for clinical practice in a rapidly advancing field. Item analysis of assessment questions is essential not only for assessing students' understanding but also for evaluating the effectiveness of the program. Thus, this study aimed to investigate the reliability (KR-20) and validity (P, D, DE) of type "A" MCQs in the Biochemistry section of the end-of-module assessment for the GIM module.

METHODS

This cross-sectional observational study was conducted at the Defence Services Medical Academy (DSMA) in Yangon, Myanmar, from April to July 2024. The end-of-module assessment was administered on June 14, 2024, from 08:00 AM to 11:00 AM. The examination comprised 100 questions, with 50% being one-best-answer (OBA) multiple-choice questions (MCQs) and the remaining 50% consisting of modified essay questions (MEQs).

The Biochemistry section included 28 OBA-type MCQs (questions 16-43), each comprising a stem with four options: one correct answer (key) and three distractors. The remaining questions assessed content from other disciplines, including Anatomy, Microbiology, Pathology, and Pharmacology. Each correct response was awarded one mark, while incorrect or unanswered responses received no marks. There was no negative marking, and the maximum possible score for the Biochemistry section was 28.

A total of 73 first-year MBBS students (30^{th} batch) participated in the end-of-module assessment for the GIM module. The inclusion criteria consisted of first-year MBBS students in 2024. The exclusion criteria included students who did not meet the required attendance percentage for the GIM module, those on sick leave or hospitalized during the examination period, and those who did not attempt all the questions in the exam.

Following the assessment, an item analysis was performed to evaluate the quality of the questions using four indices: Kuder-Richardson formula 20 (KR-20) for reliability, and the difficulty index (P), discrimination index (D), and distractor efficiency (DE) for validity. The internal consistency of dichotomous-choice exams was measured using the Kuder-Richardson 20 (KR-20) formula, a reliability coefficient suitable for dichotomous (correct/incorrect) response formats. The KR-20 score ranges from 0 to 1.0, with the following interpretation: 0.0-0.50: low reliability, 0.50-0.80: moderate reliability, and 0.80-1.00: high reliability.[11,12]

The difficulty index (P) measures how easy or difficult an item is for the students. It is calculated as the average percentage of correct responses from the top 27% and bottom 27% of performers. A score of 0.76 to 1.00 indicates an easy item, 0.25 to 0.75 indicates moderate difficulty (average), and 0.00 to 0.24 indicates a difficult item.[7-9]

The discrimination index (D) reflects how well an item differentiates between high- and low-performing students. It is the difference in correct response rates between the top and bottom 27% of students. A value of 0.40 and above indicates a very good item, 0.30-0.39 indicates a reasonably good item, 0.20-0.29 is a marginal item, and 0.19 or below is a poor item for discrimination.[7-10]

Distractor efficiency (DE) measures the effectiveness of the distractors in misleading students who do not know the correct answer. Non-functional distractors (NFDs) are those chosen by fewer than 5% of students. Items with fewer NFDs are considered to have higher distractor efficiency. Distractor efficiency ranges from 0-100% and is determined based on the number of NFDs in an item: 3 NFDs=0% DE; 2 NFDs=33.33% DE; 1 NFD=66.66% DE; no NFDs=100% DE.[7–9]

Data were entered into Microsoft Excel 2021 and analyzed using respective formulas within the Excel sheet. Results were expressed as calculated numbers and percentages. Pearson correlation analysis was used to examine the relationship between the difficulty index and the discrimination index. A p value of <0.05 was considered statistically significant. This research adhered to ethical guidelines and was approved by the ethical review committee of DSMA, Yangon, Myanmar. Since item analysis is part of routine procedures following examinations, no additional ethical concerns were identified.

RESULTS

The reliability of the biochemistry type "A" MCQs in the GIM module assessment was evaluated using the Kuder-Richardson formula 20 (KR-20), a measure of internal consistency. The KR-20 value for the test was found to be 0.682, which is below the generally accepted minimum reliability threshold of 0.8 but falls within the moderate reliability range.[11,12]

Table 1: Demographic data of the participants (n=73).

Variables		Mean±SD/ Frequency (%)
Age (years)		18.49±0.41
Gender	Male	45 (62)
	Female	28 (38)
Attendance percentage for GIM module		98.04±1.41

The results of the item analysis for the Biochemistry type "A" MCQs in the GIM module assessment are summarized based on the difficulty index, discrimination index, distractor efficiency (%), and item decision. The difficulty index categorizes the test items into three groups: easy, average, and difficult. The distribution of difficulty was as follows: easy: 7% (2 items), average: 75% (21 items), and difficult: 18% (5 items) (Table 2). The majority of the questions (75%, 21 out of 28) fell within the "average" difficulty range, suggesting that these items were appropriately challenging for the students. However, 7% (2 items) were classified as "easy," while 18% (5 items) were found to be "difficult," indicating that these items may need revision to better align with the expected difficulty level for this cohort.

Table 2: Difficulty index, discrimination index, distraction efficiency (%), and decision-making of the 28 Biochemistry MCQ questions in the GIM module.

Q No.	Difficulty index (P)		Discrimination index (D)		Distraction efficiency (%)	Decision
	(PT+PB)/N	Interpretation	(PT-PB)/n	Interpretation		
16	0.40	Average	0.50	Very good	100	Retain
19	0.45	Average	0.60	Very good	100	Retain
23	0.60	Average	0.60	Very good	100	Retain
26	0.50	Average	0.50	Very good	100	Retain
36	0.58	Average	0.75	Very good	100	Retain
37	0.45	Average	0.50	Very good	100	Retain
38	0.43	Average	0.65	Very good	100	Retain
42	0.35	Average	0.50	Very good	100	Retain
18	0.55	Average	0.80	Very good	66.66	Retain
27	0.63	Average	0.75	Very good	66.66	Retain
29	0.65	Average	0.40	Very good	66.66	Retain
32	0.58	Average	0.45	Very good	66.66	Retain
22	0.33	Average	0.35	Reasonably good	100	Retain
25	0.43	Average	0.35	Reasonably good	100	Retain
43	0.38	Average	0.35	Reasonably good	100	Retain
40	0.60	Average	0.20	Marginal item	66.66	Retain
21	0.20	Difficult	0.30	Reasonably good	100	Revise
17	0.23	Difficult	0.25	Marginal item	100	Revise
33	0.23	Difficult	0.25	Marginal item	100	Revise
20	0.28	Average	0.05	Poor item	100	Revise
34	0.38	Average	0.15	Poor item	100	Revise
35	0.48	Average	0.15	Poor item	100	Revise
39	0.33	Average	-0.05	Poor item	100	Revise
24	0.65	Average	0.00	Poor item	66.66	Revise
28	0.80	Easy	0.20	Marginal item	66.66	Revise
30	0.10	Difficult	0.10	Poor item	100	Reject
31	0.20	Difficult	0.00	Poor item	66.66	Reject
41	0.90	Easy	0.10	Poor item	33.33	Reject

PT = performance top, PB = performance bottom, N = combination of top and bottom 27% of examinees, n = 27% of examinees (top and bottom each)

The discrimination index assesses each question's ability to differentiate between high- and low-performing students. The distribution of discrimination was as follows: very good: 43% (12 items), reasonably good: 14% (4 items), marginal: 14% (4 items), and poor: 29% (8 items) (Table 2). In this analysis, 43% (12 items) showed very good discrimination, indicating that these questions effectively distinguished students of varying performance levels. However, 29% (8 items) were classified as poor items, suggesting that these questions did not adequately discriminate and may require revision or rejection.

Item decisions were as follows: retained: 57% (16 items), revised: 32% (9 items), and rejected/discarded: 11% (3 items). Based on the item analysis, 57% (16 items) were recommended for retention, indicating that they performed well in both difficulty and discrimination indices. However, 32% (9 items) required revision, and 11% (3 items) were deemed inappropriate and suggested for rejection. The following items were recommended for revision: questions 2, 5, 6, 9, 13, 18, 19, 20, and 24. Meanwhile, items 15, 16, and 26 were marked for rejection (Table 2).

The relationship between the difficulty index and discrimination index for the Biochemistry type "A" MCQs in the GIM module assessment reveals how well each question's difficulty corresponds to its ability to differentiate between high- and low-performing students. The analysis categorized the questions based on their difficulty (easy, average, difficult) and their discrimination (very good, reasonably good, marginal, poor).

Questions with very good discrimination effectively distinguished between high- and low-performing students. The majority of these items (12 questions) fell within the "average" difficulty range, indicating that they were appropriately challenging while also serving as strong discriminators. These questions were: 16, 18, 19, 23, 26, 27, 29, 32, 36, 37, 38, and 42. No items with very good discrimination were classified as easy or difficult (Table 2).

Questions with reasonably good discrimination moderately differentiated student performance. Four questions were identified in this category: average difficulty: items 22, 25, and 43; difficult: item 21. This indicates that while most of these questions were of Average difficulty, item 21 was challenging but still managed to differentiate students reasonably well (Table 2).

Questions with marginal discrimination showed limited ability to differentiate between high- and low-achievers. Four questions fell into this category: average difficulty: items 28 and 40; difficult: items 17 and 33. These results suggest that these items may need revision to improve their ability to distinguish between different levels of student understanding (Table 2).

Questions with poor discrimination did not effectively differentiate between high- and low-performing students. Eight questions fell into this category: easy: item 41; average difficulty: items 20, 24, 34, 35, and 39; difficult: items 30 and 31. Poor discrimination in both easy and difficult questions suggests that these items may either be too straightforward or too complex to serve as effective assessment tools. These questions may need to be revised or discarded (Table 2).

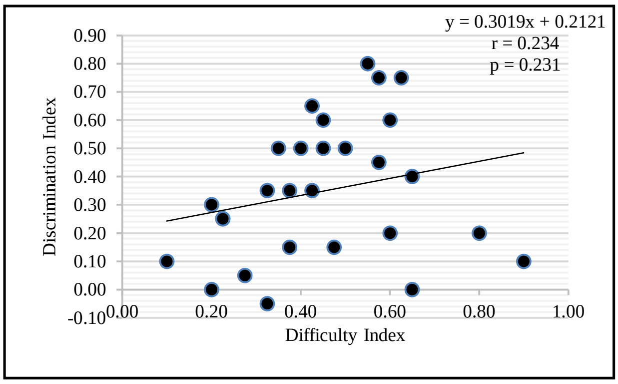

Figure 1: Correlation between difficulty index and discrimination index.

The correlation between the difficulty index and discrimination index showed a poor positive correlation (r=0.234), and this correlation was not statistically significant (p>0.05) (Figure 1). The correlation analysis indicates that most questions with good discrimination were of average difficulty. However, some questions displayed poor discrimination or were either too easy or too difficult, indicating that these items require revision or rejection to improve the overall quality of the assessment.

The distractor efficiency (DE) analysis of the Biochemistry type "A" MCQs from the GIM module highlights the quality of the distractors (incorrect options) used in each question. Non-functional distractors (NFDs) are those chosen by fewer than 5% of students, indicating that they fail to challenge students who lack a full understanding of the content.

A total of 19 question items (68%) had no NFDs, indicating that all distractors were functioning well and contributed to the challenge of the question. These questions had a distractor efficiency of 100%. Eight questions (29%) contained one NFD, which slightly reduced the distractor efficiency for these questions (66.66%). Only one question (4%) had two NFDs, with a significantly lower distractor efficiency (33.33%) (Table 3).

Table 3: Distribution of distraction efficiency (DE).

NFD	DE%	Question items	
		Frequency	Percent
0 NFD	100.00	19	68
1 NFD	66.66	8	29
2 NFD	33.33	1	4
3 NFD	00.00	0	0
Total		28	100

The following question items demonstrated the highest distractor efficiency (100%): 16, 17, 19, 20, 21, 22, 23, 25, 26, 30, 33, 34, 35, 36, 37, 38, 39, 42, and 43. Questions with moderate distractor efficiency (66.66%) had one non-functional distractor, reducing their efficiency. These items were: 18, 24, 27, 28, 29, 31, 32, and 40. Although these questions had mostly effective distractors, one option was ineffective. One question (question 41) had two NFDs, resulting in the lowest distractor efficiency (33.33%). This indicates that most of the distractors in this question failed to contribute effectively to the assessment (Table 2).

The distractor efficiency analysis shows that the majority of the questions (68%) had highly functional distractors, effectively distinguishing students with different levels of understanding. However, 29% of the questions had one NFD, and one question had two NFDs, highlighting areas where distractors may need revision to enhance the overall quality of the assessment.

DISCUSSION

This study evaluated the quality and reliability of biochemistry type "A" MCQs from the genetics, immunology, and molecular medicine (GIM) module in the year 1 MBBS program. Using the Kuder-Richardson formula 20 (KR-20), along with the difficulty index (P), discrimination index (DI), and distractor efficiency (DE), the results highlighted areas for improvement.

The mean age of the participants (first-year MBBS students) was 18.49±0.41 years, with 62% being male and the remaining 38% female. The age range of the students aligned with the admission criteria for the first-year MBBS program, with no instances of over- or under-age enrollment. The male-to-female ratio was 1.6:1, and this gender distribution did not influence the purpose or outcomes of the study (Table 1).

Students with less than 75% attendance and inadequate engagement in coursework activities- student performance, problem-based learning (PBL), presentations, and written assignments- struggled to comprehend the theoretical concepts of the module. Consequently, MCQs were more challenging, despite the fact that the questions were designed to assess essential, must-know knowledge appropriate for first-year MBBS students.[14,15] However, in this study, both attendance and coursework completion rates exceeded 96%, minimizing their potential impact on the study's findings (Table 1).

The KR-20 value of 0.682 indicates moderate reliability, suggesting that the test might not consistently measure student performance as accurately as desired. According to Brennan (2006), a KR-20 value below 0.8 signals potential reliability concerns.[11,12] To address this, regular reviews of the MCQs are necessary to ensure they are well-aligned with the learning objectives.[9] Improving the reliability of MCQs can be achieved through systematic item analysis, which evaluates test items' clarity, difficulty, and effectiveness.

Out of 28 questions, 21 (75%) had average difficulty, while 5 (18%) were classified as too difficult and required refinement. Maintaining an appropriate balance in difficulty is essential, as shown in a similar study by Shaibani et al, which emphasized the need for this balance to maintain effective evaluations of student learning.[9] The discrimination index revealed that 12 (43%) of the questions had very good discrimination, which aligns with findings from Hingorjo and Jaleel, indicating that well-constructed items tend to have high discrimination values.[7] However, 8 (29%) performed poorly, consistent with observations by Patil et al, who found that poorly discriminating items likely need revision or rejection.[16]

The DE analysis showed that 19 (68%) of the questions had fully functional distractors, a figure considered optimal for MCQ construction, as noted by Tarrant et al.[17] This result aligns with studies by Vyas and Supe, who suggested that items with three distractors perform as well as those with four, as non-functional distractors (NFDs) do not contribute to effective assessment.[18] However, 9 (32%) of the questions had at least one NFD, and one question (4%) had two NFDs, significantly reducing its effectiveness.

Based on these findings, 9 (32%) of the items were recommended for revision, and 3 (11%) for rejection. The rejected questions included those on real-time polymerase chain reaction (RT-PCR), mutation, and Down syndrome, which performed poorly in terms of difficulty, discrimination, and DE. These findings underscore the importance of revising items that do not meet the required standards of difficulty and discrimination, as supported by Velou and Ahila.[19]

Faculty development plays a key role in improving the quality of assessments. Regular training programs, such as continuing medical education (CME) and workshops, equip educators with the skills to construct high-quality MCQs that align with learning objectives and effectively assess student knowledge.[20] Faculty members trained in assessment strategies are more likely to create reliable and valid questions. Additionally, providing a variety of teaching resources, such as PowerPoint presentations, animations, and lab apparatus, can help bridge the gap between difficult and easy questions, as noted by Sood and Singh.[21]

Active learning strategies, such as problem-based learning (PBL) and group discussions, promote critical thinking and deeper understanding, which are essential for answering higher-order MCQs.[19] In contrast, relying solely on traditional lectures may limit student engagement, potentially affecting performance on assessments that test more than rote memorization.

Moreover, student learning habits and engagement with available resources play a significant role in assessment outcomes. Students who engage in independent learning, use library resources, and access online materials are more likely to succeed in assessments.[7] Those relying solely on passive learning may struggle with MCQs designed to test application and critical thinking. Addressing these factors through student support initiatives, such as promoting independent learning and resource utilization, can help level the playing field and improve outcomes.

The following recommendations are based on the findings:

Faculty should undergo regular training in assessment construction, item analysis, and MCQ development to ensure the creation of high-quality, reliable tests.

MCQs should be reviewed after each test to identify poorly performing items, with a focus on revising or rejecting those with low discrimination or non-functional distractors.

A variety of teaching and learning resources, including digital tools, demonstrations, and access to research materials, should be provided to help students engage with the content more deeply and improve their performance.

Active learning strategies, such as problem-based learning and group discussions, should be incorporated into the curriculum to foster critical thinking and prepare students for higher-order MCQs.

Promoting independent learning and ensuring that students have access to a wide range of learning resources will help them develop the skills necessary to excel in assessments.

This study underscores the importance of regular item analysis to enhance the reliability, validity, and effectiveness of MCQs in Biochemistry assessments. By addressing identified issues in question construction, discrimination, and distractor efficiency, and by improving faculty training, teaching resources, and instructional strategies, the overall quality of the assessment process can be significantly enhanced. Furthermore, promoting independent learning among students and providing them with diverse resources will contribute to better assessment outcomes, ensuring that future evaluations are both fair and reflective of student competency.

CONCLUSION

The findings of this study emphasized the importance of continuous review and revision of MCQs to enhance the reliability, validity, and overall effectiveness of assessments in medical education. Addressing key issues, such as questions with low discrimination, overly difficult or easy items, and non-functional distractors, will improve the assessment process, enabling more accurate measurement of student learning outcomes. Regular item analysis, faculty development programs, and the incorporation of active learning strategies are critical to maintaining high standards in both test design and delivery. Moving forward, ongoing efforts to improve the construction of assessments will play a vital role in fostering the academic growth and competency of medical students.

ACKNOWLEDGEMENTS

We sincerely thank the Rector of the Defence Services Medical Academy, Yangon, Myanmar, for supporting the conduct of this valuable research study. We also extend our gratitude to the Assessment Committee, Year 1 Coordinator, and GIM Module Coordinator of the Year 1 MBBS program for their invaluable contributions to this study.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Sjöström H, Christensen L, Nystrup J, Karle H. Quality assurance of medical education: lessons learned from use and analysis of the WFME global standards. Med Teach. 2019;41(6):650-5.
2. Al-Rukban MO. Guidelines for the construction of multiple choice questions tests. J Fam Community Med. 2006;13(3):125-33.
3. Black P, Wiliam D. Assessment and classroom learning. Education. 2003;20:123-33.
4. Collins J. Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radiogr Rev Publ Radiol Soc N Am Inc. 2006;26(2):543-51.
5. El-Uri FI, Malas N. Analysis of use of a single best answer format in an undergraduate medical examination. Qatar Med J. 2013;2013(1).
6. Fowell, Southgate, Bligh. Evaluating assessment: the missing link? Med Educ. 1999;33(4):276-81.
7. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012;62(2):142.
8. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore. 2006;35(2):67.
9. Shaibani TA, Ali F, Deifalla AH, Jaradat A. Item analysis of type" A" multiple choice questions from a multidisciplinary units assessment in a problem based curriculum. Bahrain Med Bull. 2021;43(2).
10. Carroll RG. Evaluation of vignette-type examination items for testing medical physiology. Adv Physiol Educ. 1993;264(6):S11.
11. Brennan RL. Educational Measurement. Rowman & Littlefield; 2023.
12. Kuder GF, Richardson MW. The theory of the estimation of test reliability. Psychometrika. 1937;2(3):151-60.
13. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005;12(1):19-24.
14. Varu M, Vegad A, Shah C, Mehta H, Kacha Y. Attendance, attitudes and academic performance: a study on first year MBBS students attending physiology classes. Int J Med Sci Educ. 2016;3(1):31-7.
15. Rasool QI, Bin SF, Farooq SU, Gani M. Exploring the correlation between attendance and academic performance in physiology among phase 1st MBBS students: a comprehensive study. Indian J Clin Anat Physiol. 2024;10(4):231-5.
16. Patil RP, Bahekar SE, Kulkarni MD, Baig MS. Evaluation of validity and reliability of multiple-choice questions in second MBBS competency-based medical education-based pharmacology examination of medical institute of India. Int J Res Med Sci. 2022;10(12):2878.
17. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in

multiple-choice questions: a descriptive analysis. BMC Med Educ. 2009;9(1):40.
18. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. Nat Med J India. 2008;21(3):130-3.
19. Velou MS, Ahila E. Refine the multiple-choice questions tool with item analysis. Int Arch Integr Med. 2020;7(8):80-5.
20. Brady AM. Assessment of learning with multiple-choice questions. Nurse Educ Pract. 2005;5(4):238-42.
21. Sood R, Singh T. Assessment in medical education: Evolving perspectives and contemporary trends. Nat Med J India. 2012;25(6):357-64.

Cite this article as: Aung YP, Soe YN, Than MM. Item analysis of type "A" multiple choice questions of biochemistry in GIM module, year 1 MBBS program. Int J Res Med Sci 2025;13:1383-90.