International Journal of Research in Medical Sciences
Htoon KZ et al. Int J Res Med Sci. 2024 May;12(5):1451-1456
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: https://dx.doi.org/10.18203/2320-6012.ijrms20241226

Original Research Article

Item analysis of multiple-choice questions in summative assessment for professional examination I of an outcome-based integrated MBBS curriculum

Kyaw Zwar Htoon[1]*, Ye Phyo Aung[2]

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

Received: 09 March 2024
Revised: 02 April 2024
Accepted: 03 April 2024

***Correspondence:**
Dr. Kyaw Zwar Htoon,
E-mail: dr.kzhtoon@gmail.com

ABSTRACT

Background: This study presents an item analysis of multiple-choice questions (MCQ-Type A) used in the summative assessment for Professional Examination I at Defence Services Medical Academy, Yangon, Myanmar. The objectives of the study were to perform item analysis using Difficulty Index (DIF I) and Discrimination Index (DI) and to correlate between DIF I and DI.
Methods: A cross-sectional observational study was conducted with 200 multiple-choice questions from two written examination papers answered by 46 medical year 2 students of Defence Services Medical Academy, Yangon, Myanmar. Item analysis of multiple-choice questions were done by using DIF I and DI calculated post-exam.
Results: Results showed that the majority of items were categorized as easy based on DIF I, with 63% and 60% in Papers I and II, respectively. Only about one-third of items were deemed acceptable, and few fell into the difficult category. DI ranged from negative to excellent, with 62% and 61% of MCQs in Papers I and II showing acceptable to excellent discrimination. Items with poor discrimination (35% and 34% in Papers I and II) should be revised or discarded. Moreover, items with negative DI should be re-evaluated for potential key errors or vague wording. A low negative correlation between DIF I and DI was observed, indicating that as DIF I increased, discrimination power decreased. Notably, items with easy DIF I demonstrated a moderate negative correlation with DI, consistent with previous research.
Conclusions: This study underscores the importance of item analysis to enhance the validity of assessment tools and ensure the effective evaluation of student cognition levels. Consequently, reconstruction and modification of MCQs are recommended to improve assessment quality and accurately measure student abilities.

Keywords: Difficulty index, Discrimination index, Item, Professional Examination 1

INTRODUCTION

Single-best-answer multiple-choice question (MCQ) or MCQ-Type A, comprises a stem, a lead-in question, and multiple options, requiring candidates to select the correct one. Despite the potential complexity in crafting them flawlessly, multiple-choice questions offer flexibility and ease of administration. They facilitate efficient answering and scoring, making them suitable for broad domain sampling and testing large number of candidates. These questions typically provide reliable testing scores per hour of testing time.[1,2] The question in MCQ or the "stem" frames the problem or scenario and provides the necessary context for respondents to formulate their

answers. Alongside the correct response, termed the "key," are the incorrect options, which are referred to as "distractors." These distractors are intentionally crafted to attract students who may not have grasped the material fully or are uncertain about the correct solution. Collectively, the stem, distractors, and key comprise the entire MCQ, commonly referred to as the "item".[3] Item analysis is regularly used for assessing the quality of MCQs in an examination. It is a useful and straightforward procedure carried out after an examination. It helps to determine how reliable and valid a test item is. Moreover, item analysis can also assist instructors in improving their ability to create tests and identifying which parts of the course content need more attention or explanation.[4] Item analysis of MCQs could find out which MCQs are defective and which of those are needed to revised or removed from the question bank. The quality of an MCQ item could be determined by three indices, Difficulty Index (DIF I), Discrimination Index (DI) and Distractor Efficiency (DE).[5–7]

Defence Services Medical Academy has changed its MBBS curriculum from the traditional discipline-based curriculum to outcome-based integrated curriculum since 2017. In Phase I of MBBS curriculum, the students have to pass two academic years in which twelve system modules (Basic Medical Science Module, Basic Mechanism of Disease Module, Genetics, Immunology and Molecular Medicine Module, Cardiovascular System Module, Hematopoietic System Module, Respiratory System Module, Nervous System Module, Skin and Musculoskeletal System Module, Urinary System Module, Gastrointestinal System and Nutrition Module, Reproductive System Module, Endocrine System and Metabolism Module), allotted in four semesters. Every academic year consists of 37 weeks of teaching-learning and each module extends from 4 to 6 weeks. At the end of medical year 2, the students must sit for summative assessment for Professional Examination I. In this assessment, two examinations papers were conducted to assess the medical year 2 students' quality. Each student must pass two written papers and two sets of OSCE/OSPE examination. Each written paper contains two parts: 100 single-best-answer MCQs in Part A and 6 modified essay questions in Part B. In this study, 200 single-best-answer MCQs from written Paper I and II of summative assessment for Professional Examination I of Defence Services Medical Academy, Yangon, Myanmar is analysed by using DIF I and DI.

This study aimed to carry out item analysis of single-best-answer MCQs used in summative assessment of Professional Examination I by using DIF I and DI. Also, to correlate DIF I and DI of the studied single-best-answer MCQs.

METHODS

This study was a cross-sectional observational study conducted from October 2023 to December 2023 at Defence Services Medical Academy, Yangon, Myanmar. A total of 46 undergraduate medical students of medical year 2 were assessed in summative assessment for professional examination I in November 2023. The students had to answer total 200 single-best-answer MCQs in written Paper I and Paper II of summative assessment for professional examination I.

Each item comprised a stem along with four options, among which one was the key, while the remaining three served as distracters. A scoring system was employed where a correct response earned a score of '1', while an incorrect answer or no response resulted in a score of '0', with no penalty for incorrect choices.

Following the examination, item analysis was conducted using DIF I and DI. The analysis was performed at the Department of Medical Education of Defence Services Medical Academy, Yangon, Myanmar.

For item analysis, the results of each examination paper were arranged in descending order from highest marks to lowest marks. Then, they were divided in three groups. The upper third (15) and lower third (15) were included in study and designated as high scoring group (H) and low scoring group (L) respectively. The middle one third (16) scoring medium marks were excluded, assuming they are in average. Each item was analysed for DIF I and DI. DIF I and DI were calculated as described in other studies.[8–12] The formulas used in calculation are as follows:

$$DIF\ I = [(H+L) / N] \times 100$$

where; H= number of students choosing correct answer in high score group, L= number of students choosing correct answer in low score group, N= total number of students in both groups.

DIF I of an item range between 0-100%. DIF I is categorized as: 1) DIF I >70% = Easy, 2) DIF I between 30-70% = Acceptable, and DIF I <30 = Difficult.

DI is the ability of a MCQ to differentiate the students getting high scores from low scoring ones. Formula used to calculate DI is

$$DI = 2 \times [(H-L)/N]$$

DI is categorized as, 1) DI Negative = Defective item/wrong key, DI between 0-0.19 =Poor discrimination, DI between 0.2-0.29 = Acceptable discrimination, DI between 0.3-0.39 = Good discrimination, DI ≥0.4 = Excellent discrimination.

The data were inserted into Microsoft Excel 2021 and were analysed by using SPSS software version 25. The data were reported as % and mean ± standard deviation of all items. The relationship between the item difficulty index and DI values for all items was determined using

Pearson correlation analysis. P value <0.05 was used to indicate statistical significance.

The present research work did not contain any studies performed on animals/human subjects by any of the authors and was approved by the Ethical Review Committee of Defence Services Medical Academy, Yangon, Myanmar as item analysis is the routine procedure following every examination.

RESULTS

There were 100 items in each written paper and the questions were constructed by Departments of Anatomy, Physiology, Biochemistry, Microbiology, Pathology and Pharmacology of Defence Services Medical Academy. The distribution of the items for the written Paper I and II constructed by each department was described in Table 1.

Table 1: Distribution of the items constructed by related departments for medical year 2 for Professional Examination 1.

Name of Department	Number of items	
	Written Paper I	Written Paper II
Anatomy	23	15
Physiology	16	20
Biochemistry	20	21
Microbiology	12	11
Pathology	14	18
Pharmacology	15	15
Total	100	100

Mean difficulty indices (DIF I) of the items in written Paper I and Paper II were 73.18±18.36 and 72.73±23.34 respectively. Mean Discrimination Indices (DI) of written Paper I and Paper II were 0.26±0.18 and 0.25±0.23 respectively (Table 2).

Table 2: Parameters of the items in written paper I and II for summative assessment for Professional Examination 1.

Examination paper	Written Paper I	Written Paper II
No. of items	100	100
No. of students	46	46
Difficulty index (DIF I) %		
Mean±SD	73.18±18.36	72.73±23.34
Range	21.43-96.43	6.67-100
Discrimination index (DI)		
Mean±SD	0.26±0.18	0.25±0.23
Range	-0.14-0.79	-0.40-0.80

Majority of the MCQs in Paper I and Paper II were in easy category (DIF I > 70), 63 and 60 items out of 100 in each paper. In acceptable category (DIF I=30-70), 36 and 33 items of the written Paper I and Paper II were included. One item from Paper I and 7 items from Paper II fell into the difficult category (DIF I < 30) (Table 3).

Table 3: Distribution of DIF I of single-best-answer MCQs in written paper I and II.

IDF 1	Written Paper I N (%) (n=100)	Written Paper II N (%) (n=100)	Interpre-tation	Action
<30	1	7	Difficult	Revise/ discard
30-70	36	33	Acceptable	Store
>70	63	60	Easy	Revise/ discard

Nearly one-fourth of the all items, 24 in Paper I and 25 in Paper II were in excellent discrimination range (DI >0.4), while 7 and 14 items from the two Papers were in good discrimination range (DI=0.3-0.39). About one-third of the items (31 items) in Paper I and one-fifth of the items (22 items) in Paper II were in acceptable discrimination range (DI=0.2-0.29). There were 3 items of Paper I and 5 items from Paper II were in negative discrimination, while 35 and 34 items of Paper I and Paper II were in poor discrimination range (DI=0-0.19) (Table 4).

Table 4: Distribution of DI of single-best-answer MCQs in written paper I and II.

DI	Paper I N (%) (n=100)	Paper II N (%) (n=100)	Interpre-tation	Action
>0.4	24	25	Excellent	Store
0.3-0.39	7	14	Good	Store
0.2-0.29	31	22	Acceptable	Store
0-0.19	35	34	Poor	Revise/ discard
Negative	3	5	Defective item/ wrong key	Revise/ discard

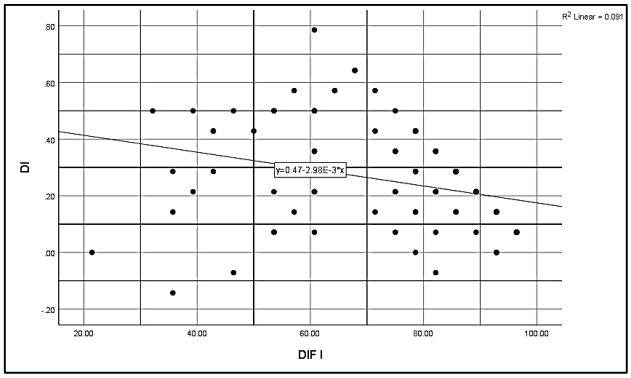

Figure 1: Scatter plot showing the relationship between DIF I and DI of the single-best-answer MCQs in written paper I (n=100).

There was low negative correlation between DIF I and DI of paper I (r=-0.301, p=0.02) (Figure 1) and paper II (r=-0.344, p=0.000) (Figure 2). Sixty-three MCQs of easy DIF I category of Paper I was also moderately negatively correlated (r=-0.607, p=0.000) with their relative DI (Figure 3). Similarly, 60 MCQs of easy DIF I category of Paper II had moderate negative correlation with their DI (r=-0.669, p=0.000) (Figure 4).

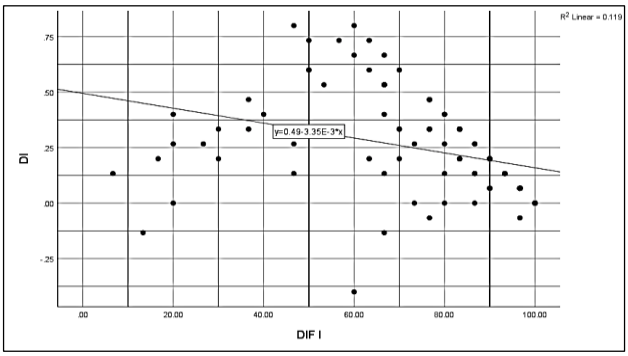

Figure 2: Scatter plot showing the relationship between DIF I of the single-best-answer MCQs in written paper II (n=100).

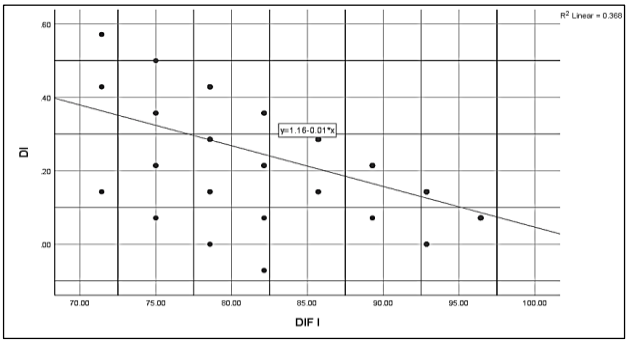

Figure 3: Scatter plot showing the relationship between DIF I and DI of the single-best-answer MCQs with easy difficulty index in written paper I (n=63).

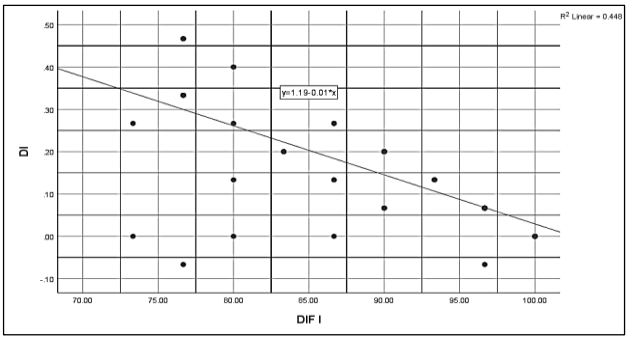

Figure 4: Scatter plot showing the relationship between DIF I and DI of the single-best-answer MCQs with easy difficulty index in written paper II (n=60).

Low positive correlation (r=0.322, p=0.056) was seen in 36 MCQs of Paper I between acceptable DIF I range and their respective DI (Figure 5). Thirty-three MCQs of acceptable DIF I range in Paper II were also low positive correlation (r=0.035, p=0.847) with DI (Figure 6).

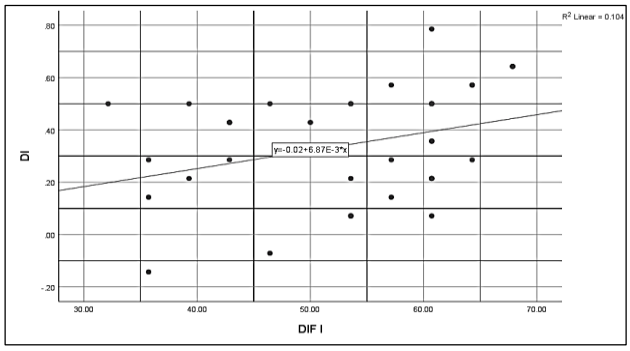

Figure 5: Scatter plot showing the relationship between DIF I and DI of the single-best-answer MCQs with acceptable difficulty index in written paper I (n=36).

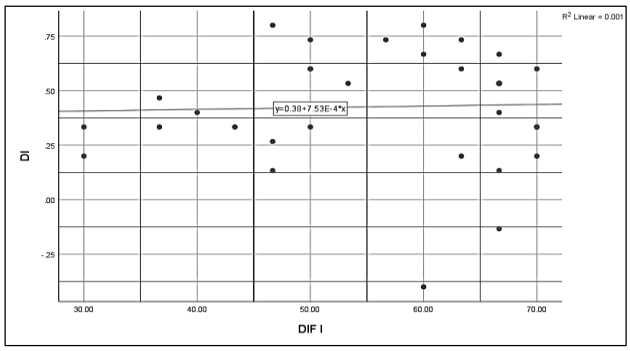

Figure 6: Scatter plot showing the relationship between DIF I and DI of the single-best-answer MCQs with acceptable difficulty index in written paper II (n=33).

DISCUSSION

Assessment offers students feedback on their progress, fostering deeper engagement with the material and effective learning strategies. Furthermore, aligning curriculum goals with learning outcomes through assessment ensures students master the intended skills and knowledge by matching activities and evaluations accordingly.[13] Many assessment tools such as multiple-choice questions, short answer questions, modified essay questions etc. are being utilized to assess the level of cognition as per Bloom's taxonomy in either formative or summative assessment in medical universities. Defence Services Medical Academy currently uses single-best-answer MCQs and modified essay question in assessment of the students.

Single-best-answer MCQ allows for recognition of consonance between two facts, attributes, or concepts, is easier to write than true/false questions, and permits more subtle distinctions in knowledge. However, item analysis

should be made after every assessment to ensure the quality of the MCQs.[2] This present study was done to analyse the single-best-answer MCQ items of summative assessment for Professional Examination 1 in Defence Services Medical Academy by studying DIF I and DI.

Majority of the items in this study were in easy category of the DIF I, and 63% and 60% of the Paper I and II were easy. Only about one-thirds of all the items in both papers were acceptable and few items were in difficult category. The percentage of easy items in this research was higher than other studies done by Teli and Kate, Shete et al, Date et al, Kausar et al and Baste.[5,8-10,14] In these studies, easy items were within 20 to 30% of total and acceptable ones were within the range of 28 to 70% of all items. However, numbers of items in these analyses were small and only between 40 and 64. Easy and difficult items need to be evaluated and reconstructed to be able to be used in next examination. Easy items could also be used in formative assessment because they help to boost students' confidence. There should be few difficult items in an examination paper as they help to rank the high achievers from average students. However, the examination paper should be constructed with the majority of acceptable questions per DIF I.

Majority of MCQs, 62% and 61% in Paper I and II, were acceptable to excellent in terms of DI, and they should be kept in question bank for future summative assessment. There were items with poor discrimination power (35% and 34% of Paper I and II) and they should be revised or discarded from the question bank. There were 3 and 5 items which showed negative DI value in Paper I and II. Items with negative DI in the present study were fewer than the study by Yahia in which 8.4% of the questions were with negative DI values and the study by Gajjar et al in which 20% were with negative DI.[15,16] A negative DI may be due to the wrong key or vague wordings in the questions. These items should be revised and reframed to be useful for future assessment.

This study showed a low negative correlation between difficulty and DI. It was different from the studies by Shete et al and Kausar et al in which DIF I correlated positively with DI.[8,10] The difficulty index is also known as the ease index because the more DIF I number increases, the easier the question is.[11] Therefore, when the DIF I increase, the discrimination power of an item should be lower. Discrimination was noted to be maximum in the upper range of acceptable difficulty (30% and 80%) but lower in easy and difficult items in both examination papers. Items with easy DIF I category of both examination papers showed moderate negative correlation with DI and these correlations were statistically significant. This finding was consistent with the result of the study by Kausar et al.[10] There were low positive correlations MCQs of Paper I between acceptable DIF I range and their respective DI in both examination papers. However, these data were not statistically significant.

This study has some limitations. Distractor Efficiency (DE) should be also analysed for the quality assessment of the items. However, the total distractors were 600 in two examination papers and the analysis of distractor efficiency for all items would be time consuming. There would be non-functional distractors which should be rechecked and modified in every item of easy difficulty index. As they would be reconstructed for the next examination, the authors decided not to analyse the distractor efficiency for the items in this study.

CONCLUSION

Assessment of single-best-answer MCQs from written Paper I and Paper II of summative assessment for Professional Examination I in Defence Services Medical Academy highlights the importance of item analysis after examination. In this study, large numbers of MCQs were easy but good in discrimination to distinguish higher ability (high achiever) students and lower ability (low achiever) students. Majority of the single-best-answer MCQs from Paper I and Paper II must be thoroughly reconstructed and modified in order to improve the validity of the assessment and to be able to assess the level of cognition of the students.

ACKNOWLEDGEMENTS

Authors would like to thank the teachers of Department of Anatomy, Physiology, Biochemistry, Microbiology, Pathology and Pharmacology of Defence Services Medical Academy for constructing MCQs and their efforts in summative assessment for Professional Examination 1.

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

REFERENCES

1. Schuwirth LWT, van der Vleuten CPM. Written assessment. In: Dent JA, Harden RM, editors. A practical guide for medical teacher. 4th ed. Churchill Livingstone; 2013:299-306.
2. Jolly B. Written assessment. In: Swanwick T, editor. Understanding medical education: evidence, theory, and practice. 2nd ed. Wiley Blackwell; 2014:255-304.
3. Al-Rukban MO. Guidelines for construction of multiple choice questions tests. J Fami Commu Med. 2006;13(3):125-133.
4. Collins J. Writing multiple choice questions for continuing medical education activities and self-assessment modules. Radiogra. 2006;26(2):543-1.
5. Teli C, Kate N. Item analysis of multiple-choice questions in anatomy for first year MBBS. Natl J Physiol Pharm Pharmacol. 2022;12(3):1529-32.

6. Suryadevara VK, Bano Z. Item analysis to identify quality multiple choice questions/items in an assessment in Pharmacology of II MBBS students in Guntur Medical College of Andhra Pradesh, India. Int J Basic Clin Pharmacol. 2018;7(8):1517-21.
7. Namdeo S, Sahoo B. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. Int J Res Med Sci. 2016;1716-9.
8. Shete A, Kausar A, Lakhkar K, Khan S. Item analysis: An evaluation of multiple choice questions in physiology examination. J Contemp Med Educ. 2015;3(3):106-9.
9. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. Int J Basic Clin Pharmacol. 2019;8(9):1999-2003.
10. Kausar A, Daimi S, Borulkar T. Assessing an assessment tool: Analysis of multiple choice questions on difficulty level and discrimination power, from an assessment in physiology. Natl J Physiol Pharm Pharmacol. 2022;(06):801-5.
11. Chauhan GR, Chauhan BR, Vaza J V, Chauhan PR. Relations of the number of functioning distractors with the item difficulty index and the item discrimination power in the multiple choice questions. Cureus. 2023;15(7).
12. Al Ameer AY. Assessment of the quality of multiple-choice questions in the surgery course for an integrated curriculum, University of Bisha College of Medicine, Saudi Arabia. Cureus. 2023;15(12).
13. McKimm J, Forrest K, Thistlethwaite J. Medical education at a glance. 1st ed. Wiley Blackwell; 2017.
14. Baste V. Item analysis of MCQs in physiology and its correlation with faculty's perception of difficulty level of MCQs. Natl J Physiol Pharm Pharmacol. 2023;13(7):1444-7.
15. Yahia AIO. Post-validation item analysis to assess the validity and reliability of multiple-choice questions at a medical college with an innovative curriculum. Natl Med J India. 2021;34(6):359-62.
16. 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. Ind J Commu Med. 2014;39(1):17-20.

Cite this article as: 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;12:1451-6.