DOI: https://doi.org/10.4038/seajme.v19i2.687

Original Research Paper

Effect of Option Number on Psychometric Quality of Single-answer MCQs in Medical Exams

Khan, S.A., Das, B.K.L., Lamsal, M.

Abstract

Introduction: Multiple choice questions (MCQs) allow efficient student assessment and broader content coverage in less time. However, crafting clear stems and plausible options is challenging and time-consuming. Despite research supporting three-option MCQs, many medical educators continue using four or five options due to limited evidence from undergraduate medical settings. Therefore, this study aimed to compare the quality indices of MCQs with 3, 4, and 5 options through item analysis and to determine the optimal number of options for effective assessment.

Methods: The study included 107 first-year and 91 second-year MBBS students. Tests were conducted two weeks after internal assessments. Students were randomly divided into three groups: first year (36, 35, 36) and second year (29, 31, 31). Three sets of 30 biochemistry MCQs were used, each with 3, 4, or 5 options for years. For every item, the Difficulty Index (DIF I), Discrimination Index (DI), and Distractor Effectiveness (DE) were analyzed.

Results: In 1st year MBBS, 3-option MCQs were easiest (DIF I 69.28%) with highest distractor effectiveness (83.33%) and moderate discrimination (0.27). Four-option MCQs showed the highest discrimination (0.30) but lower distractor effectiveness (69.96%). Five-option MCQs were hardest (DIF I 57.36%) but had lowest discrimination (0.21) and distractor effectiveness (64.16%). In 2nd year, 3-option MCQs had highest difficulty (53.66%) and discrimination (0.32), while 4-option MCQs had best distractor effectiveness (86.64%). Non-functional distractors increased with more options.

Conclusion: Reducing MCQ options in Biochemistry to three improves ease and distractor functionality without compromising discrimination, making them a more efficient and effective choice for assessment design.

Keywords: Psychometric indices, Difficulty index, Discrimination index, Distractor effectiveness

Introduction

A multiple choice question (MCQ) is a type of written assessment consisting of a scenario (stem) followed by a question known as a lead in and three or more options from which examinees need to choose the correct response. The rest of the options are known as distractors (Cizek & O'Day, 1994).

Department of Biochemistry, B.P. Koirala Institute of Health Sciences, Nepal

Corresponding author: Dr. Seraj Ahmed Khan
Email: drserajkhan@gmail.com

The MCQ format allows the examiners to assess large number of students with a wide range of content in less given time (Downing, 2002; McCoubrie, 2004). The popularity of MCQ is because it allows for diagnostic as well as formative purpose. Scoring is easy and subject wise bias is almost absent (Bruno & Dirkzwager, 1995). Some researchers agree on the fact that quality of distractors of MCQs is more important than its number. Researchers have also found that it is very difficult to write three or four plausible distractors for each item. They recommend decreasing the number of choices per item but making the test longer so that it can cover

more content area in the given time frame and better discriminate among high-performing students (Haladyna & Downing, 1993).

Earlier Owen and Froman (1987) have advocated the advantages of 3 options MCQs over 5 options. However, examiners still continued to advocate the use of four or five options. More recent studies have also questioned the standard use of four or five options, finding three options sufficient in most cases (Dehnad *et al.,* 2014; Vegada *et al.,* 2016; Farhady & Shakery, 2000; Vyas & Supe, 2008; Tarrant & Ware, 2010; Lee & Winke, 2013). In a meta-analysis of 80 years multiple choice questions, Rodriguez (2005) reported that three option tests helped to include vast content in less time at without affecting the psychometric quality of the tests.

Recent literature has reinforced the advantages of 3 option MCQs, however many medical schools around the world, including ours, still follow the conventional use of 4 or 5 option MCQs for objective assessment or in entrance examinations. Therefore, this study was an attempt to examine the quality indices of 3, 4 and 5 option MCQs in Biochemistry and provide more evidence to reinforce the usefulness of 3 options over 4 and 5 options when constructing MCQs.

Methods

Study Design and Participants

The study included 107 first-year and 91 second-year MBBS students. Written informed consent was obtained from all participants prior to enrolment, and they were informed about the test 10 days in advance. The test was conducted two weeks after the Unit 0 (Basic Concepts) internal assessment for first-year students and the Unit 3 (GIT, Hepatobiliary, Nutrition) assessment for second-year students.

Test tool

Both first and second year MBBS students, were tested with thirty standardized five option MCQs in Biochemistry, selected from the departmental question bank, based on Units 0 and 3. From these, two additional sets were created by systematically removing one option for a four-option set, and two options for a three-option set. Thus, set 1 contained three-option MCQs, set 2 had four-option MCQs, and Set 3 included the original five-option MCQs. The MCQ stems and the order of distractors remained consistent across all sets. The questions were reviewed and validated by subject experts, and their feedback was incorporated. A total of 45 minutes was allotted for the test.

Objective Assessment

Separate days were allocated for conducting the test for first and second year MBBS students. To minimize bias, students were randomly divided into three groups and seated in separate classrooms. In the first year, 107 students were divided into groups of 36, 35, and 36. In the second year, 91 students were divided into groups of 29, 31, and 31. Each group received a different MCQ set: Group 1 received Set 1 (three-option), Group 2 received Set 2 (four-option), and Group 3 received Set 3 (five-option). The test duration was explained in advance, and students were instructed to complete the test within the allotted time.

Calculation of Indices

- Difficulty index (DIF I) or p value was calculated using formula, $p = [(H + L)/N] \times 100$

  H = Number of students answering the item correctly in the high achieving group

  L = Number of students answering the item correctly in the low achieving group

  N = Total number of students in the two groups (including non-responders)

- Discrimination index (DI) or d value was calculated using formula, $d = 2 \times [(H - L)/N]$

  Where the symbols H, L, and N represent the same values as mentioned above.

- Distractor effectiveness (DE) or functionality: DE was calculated on the basis of Non-functional distractor (NFD). NFD in an item is option (s) (other than key) selected by <5% of students. Alternatively functional distractors are those selected by 5% or more students.DE is determined for each item on the basis of the number of NFDs in it and ranges from 0 to 100%. If an item contains three or two or one or nil NFDs then DE will be 0, 33.3, 66.6, and 100%, respectively. The higher the DE the more effective the MCQ is.

Table 1: Interpretation of Difficulty index and Discriminating index

Difficulty index		Discriminating index	
P Score	**Interpretation**	**D Score**	**Interpretation**
<30%	Difficult	≥0.35	Excellent
50-60%	Ideal	0.25-0.34	Good
30-50% or 60%-70%	Acceptable	0.21-024	Acceptable
>70%	Very easy	≤0.20	Poor

Statistical Analysis

The scores of the students was entered in Microsoft Excel and arranged in descending order to identify the upper 1/3 higher achieving group and last 1/3 lower achieving group. The data were presented in mean±SD, frequency and percentage. Normality of the data was tested by Kolmogorov-Smirnov test. DIF I and DI was calculated as per the standard formulas in Microsoft Office Excel 2013. DE was calculated on the basis of number of Non Functional Distractor (NFD). Correlation analysis was done by Pearson's coefficient. One-way analysis of variance (ANOVA) tests were carried out to compare the differences in mean scores between 3 sets MCQs. Analysis was done by SPSS 20.0 (IBM, Armonk, NY, United States of America). $p<0.05$ was considered to indicate statistical significance.

Results

The performance of multiple-choice questions (MCQs) with 3, 4, and 5 options was evaluated across three psychometric parameters: Difficulty Index (DIF I), Discrimination Index (DI), and Distractor Effectiveness (DE) in 1st and 2nd year MBBS students.

Table 2 summarizes the comparison of psychometric indices among the three sets of MCQs in the 1st year cohort.

Difficulty Index (DIF I):
The 3-option MCQs (Set 1) had the highest mean difficulty index (69.28%), followed by 4-option (67.91%) and 5-option MCQs (57.36%). This suggests that 3-option MCQs were relatively easier for students.

Discrimination Index (DI):
The highest DI was seen in 4-option MCQs (0.30), followed by 3-option (0.27) and 5-option (0.21) MCQs. All sets fell within the acceptable to good range of discrimination, with 5-option MCQs having the lowest ability to differentiate between high and low performers.

Distractor Effectiveness (DE):
3-option MCQs showed the highest distractor effectiveness (83.33%), indicating more functional distractors. In contrast, DE was lower in 4-option (69.96%) and lowest in 5-option MCQs (64.16%).

Correlation Analysis:
A weak negative correlation between DI and DIF I was found ($r=–0.158$, $p=0.136$), which was statistically non-significant (Figure 1).

Table 2: Comparison of different indices among the 3 sets of MCQ's in MBBS 1st Year

Indices	Set 1 (3 options)	Set 2 (4 options)	Set 3 (5 options)	P value*
DIF I	69.28%	67.91%	57.36%	
DI	0.27	0.30	0.21	
DE	83.33%	69.96%	64.16%	

**One-way ANOVA was used to compare the mean across the groups. p value <0.05 was considered to be statistically significant; DIF I: Difficulty Index; DI: Discrimination Index; DE: Distractor Effectiveness*

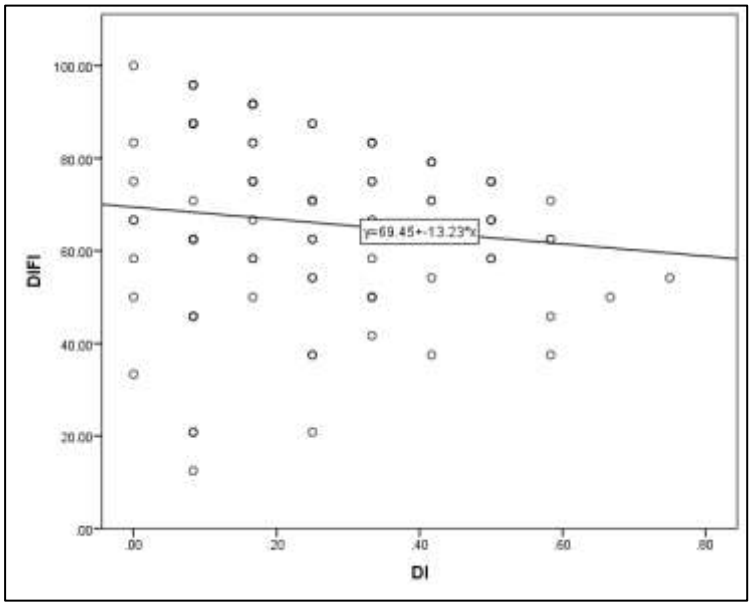

Figure 1: Correlation between DI and DIF I of the given items in MBBS 1st Year. (r= -0.158, p=0.136)

The psychometric comparison in 2nd year students is shown in Table 3.

Difficulty Index (DIF I):
The 3-option MCQs (Set 1) again showed the highest DIF I (53.66%) followed by 5-option (44.4%) and 4-option (40.8%) MCQs, suggesting moderate to acceptable levels of difficulty across all sets.

Discrimination Index (DI):
The DI was fairly consistent across the three sets: 3-option (0.32), 4-option (0.30), and 5-option (0.30), all falling within the good discrimination range.

Distractor Effectiveness (DE):
The highest DE was found in 4-option MCQs (86.64%), closely followed by 3-option (83.33%), while 5-option MCQs had a markedly lower DE (54.17%). There were 10, 27, 43 Non-functional distractors in the Set 1, 2 and 3 respectively.

Correlation Analysis:
A weak and non-significant negative correlation between DI and DIF I was noted (r=–0.128, *p*=0.230) (Figure 2).

Table 3: Comparison of different indices among the 3 sets of MCQs in MBBS 2nd Year

Indices	Set 1 (3 options)	Set 2 (4 options)	Set 3 (5 options)	P value*
DIF I	53.66%	40.8%	44.4%	
DI	0.32	0.30	0.3	
DE	83.33%	86.64%	54.17	

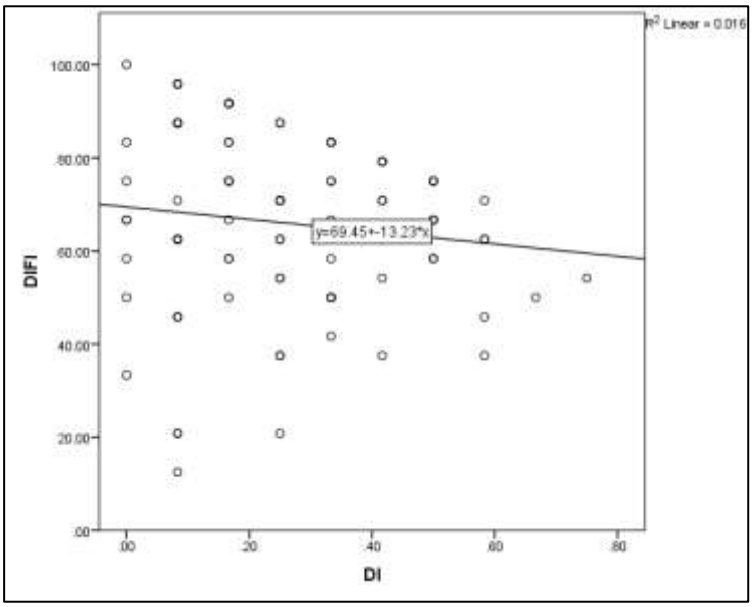

Figure 2: Correlation between DI and DIF I of the given items in MBBS 2nd Year. (r= -0.128, p=0.230)

Discussion

The present study evaluated the effect of the number of options in single-answer multiple choice questions (MCQs) in Biochemistry on key psychometric parameters—difficulty index (DIF I), discrimination index (DI), and distractor effectiveness (DE) among undergraduate medical students. The findings revealed that 3-option MCQs performed either equally well or better across most psychometric indices when compared to 4 and 5-option formats.

The difficulty index was highest for 3-option MCQs in both 1st and 2nd year students, indicating that these items were easier to attempt correctly. Previous research suggests that ease of an item does not necessarily compromise its ability to discriminate between high- and low-performing students, especially when difficulty is within an ideal range of 50–60% (Bruno & Dirkzwager, 1995; Haladyna & Downing, 1993). In our study, the discrimination index for 3-option MCQs was also within the acceptable to good range, consistent with findings by Rodriguez, who concluded that reducing the number of options to three did not negatively affect test discrimination (Rodriguez, 2005).

The most notable finding in our study was the superior distractor effectiveness observed in 3-option MCQs. A high proportion of functional distractors indicates better item quality, as non-functional distractors those selected by less than 5% of students tend to reduce item reliability and add unnecessary complexity (Tarrant & Ware, 2010). Our results are in line with those of Tarrant and Ware, who found that items with three plausible options often outperformed items with four or five, particularly in terms of functional distractors (Tarrant & Ware, 2008). In contrast, 5-option MCQs in our study had the lowest distractor effectiveness, suggesting that the additional options were often implausible or misleading.

The slightly better discrimination index seen in 4-option MCQs among 1st year students (0.30) compared to 3-option MCQs (0.27) was not statistically significant, and the trade-off came with reduced distractor functionality. This supports earlier claims that the quality of distractors is more critical than their number (Owen & Froman, 1987). Writing plausible distractors is a well-documented challenge, and attempts to include more options often result in non-functional choices, lowering the

overall validity of the question (Scholz, 2016; Crehan *et al.,* 1993)

Correlation analysis between difficulty and discrimination indices showed a weak, negative, and statistically non-significant relationship, suggesting that moderately difficult questions were just as likely to distinguish between high- and low-performing students. Similar findings have been reported by Haladyna and Downing, emphasizing that discrimination is optimized when item difficulty is neither too low nor too high (Haladyna & Downing, 1989).

Despite accumulating evidence in favor of the 3-option MCQ format, many medical schools, including ours, continue to use traditional 4- and 5-option formats in internal and competitive examinations. This persistence may be due to historical conventions or misconceptions about test reliability. However, several meta-analyses have confirmed that tests using 3-option MCQs maintain equivalent reliability and validity while allowing for broader content coverage and reduced cognitive burden for students (Rodriguez, 2005; Attali & Bar-Hillel, 2003).

Our findings provide further support for revisiting and revising MCQ construction policies, especially in resource-limited settings, where efficient and valid assessment methods are needed. Given the improved distractor functionality and acceptable difficulty and discrimination levels, the 3-option format appears to be both pedagogically and psychometrically superior.

Conclusion

This study concludes that 3-option MCQs offer several advantages over 4- and 5-option MCQs in undergraduate medical assessments in Biochemistry. They demonstrate higher or comparable difficulty and discrimination indices, and notably better distractor effectiveness. Given the challenges in developing multiple plausible distractors, using three well-constructed options can enhance test quality and efficiency without compromising validity or reliability.

Medical educators and institutions should consider adopting the 3-option MCQ format for routine assessments and entrance examinations to promote better assessment practices.

Acknowledgements

Authors wish to acknowledge the BPKIHS medical undergraduate students who participated in this study and to Dr. Bishal Raj Joshi, Saroj Kunwar and Suman Sapkota who helped in organizing and conducting the test.

Availability of data and materials

The data used and analyzed during the current study has been included in the results section. However, data can be made available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Review Committee of B.P. Koirala Institute of Health Sciences, Dharan, Nepal. Code of the approval letter was IRC/464/015. Informed consent was taken from the students before including them in the study.

References

Attali, Y. & Bar-Hillel, M. (2003) 'The position of correct answers in MCQs as a psychometric variable', *Journal of Educational Measurement*, 40(2), pp. 109–128.

Bruno, J.E. and Dirkzwager, A. (1995). Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educational and Psychological Measurement*, 55(6), pp.959-966.

Cizek, G.J. & O'Day, D.M. (1994) 'Further investigations of nonfunctioning options in multiple-choice test items', *Educational and Psychological Measurement*, 54, pp. 861–872.

Crehan, K.D., Haladyna, T.M. & Brewer, B. (1993) 'Use of an inclusive option and optimal number of MCQ options', *Educational and Psychological Measurement*, 53(2), pp. 241–247.

Dehnad, A., Nasser, H. & Hosseini, A.F. (2014) 'A comparison between three- and four-option multiple choice questions', *Procedia – Social and Behavioral Sciences*, 98, pp. 398–403.

Downing, S.M. (2002) 'Assessment of knowledge with written test forms', in Norman, G.R., Van der Vleuten, C. & Newble, D.I. (eds.) *International Handbook of Research in Medical Education*. Dordrecht: Kluwer Academic Publishers, pp. 647–672.

Farhady, H. & Shakery, S. (2000) 'Number of options and economy of multiple-choice tests', *Roshd Foreign Language Journal*, 14(57).

Haladyna, T.M. & Downing, S.M. (1989) 'Validity of a taxonomy of MCQ item-writing rules', *Applied Measurement in Education*, 2(1), pp. 51–78.

Haladyna, T.M. & Downing, S.M. (1993) 'How many options are enough for a multiple-choice test item?', *Educational and Psychological Measurement*, 53(4), pp. 999–1010.

Lee, H. & Winke, P. (2013) 'Differences among 3-, 4-, and 5-option items in a high-stakes English test', *Language Testing*, 30(1), pp. 99–123.

McCoubrie, P. (2004) 'Improving the fairness of multiple-choice questions: a literature review', *Medical Teacher*, 26, pp. 709–712.

Owen, S.V. & Froman, R.D. (1987) 'What's wrong with three-option multiple choice items?', *Educational and Psychological Measurement*, 47, pp. 513–522.

Rodriguez, M.C. (2005) 'Three-options are optimal for multiple-choice items: A meta-analysis', *Educational Measurement: Issues and Practice*, 24(2), pp. 3–13.

Scholz, C. (2016) 'Number of answer options in MCQs: A meta-analysis', *Medical Teacher*, 38(2), pp. 123–129.

Tarrant, M. & Ware, J. (2008) 'Impact of item-writing flaws in MCQs on achievement', *Medical Education*, 42(2), pp. 198–206.

Tarrant, M. & Ware, J. (2010) 'A comparison of psychometric properties of three- and four-option MCQs', *Nurse Education Today*, 30(6), pp. 539–543.

Vegada, B., Shukla, A., Khilnani, A. et al. (2016) 'Comparison between three option, four option and five option MCQ tests: a randomized study', *Indian Journal of Pharmacology*, 48(5), p. 571.

Vyas, R. & Supe, A. (2008) 'Multiple choice questions: a literature review on the optimal number of the options', *National Medical Journal of India*, 21(3), pp. 130–133.