International Journal of Scientific Reports
Alemu AT et al. Int J Sci Rep. 2024 Jun;10(6):195-199
http://www.sci-rep.com
pISSN 2454-2156 | eISSN 2454-2164
DOI: https://dx.doi.org/10.18203/issn.2454-2156.IntJSciRep20241316

Original Research Article

Quality of multiple choice question items: item analysis

Ayenew Takele Alemu*, Hiwot Tesfa, Addisu Mulugeta, Enyew Tale Fenta, Mahider Awoke Belay

Department of Public Health, College of Medicine and Health Sciences, Injibara University, Injibara, Ethiopia

Received: 08 March 2024
Accepted: 11 April 2024

***Correspondence:**
Ayenew Takele Alemu,
E-mail: ayetake21@gmail.com

ABSTRACT

Background: There are different types of exam formats for educational assessment. Multiple choice questions (MCQs) are frequently utilized assessment tools in health education. Considering the reliability and validity in developing MCQ items is vital. Educators often face the difficulty of developing credible distractors in MCQ items. Poorly constructed MCQ items make an exam easier or too difficult to be answered correctly by students as intended learning objectives. Checking the quality of MCQ items is overlooked and too little is known about it. Therefore, this study aimed to assess the quality of MCQ items using the item response theory model.
Methods: A descriptive cross-sectional study was conducted among MCQ items of public health courses administered to 2nd year nursing students at Injibara university. A total of 50 MCQ items and 200 alternatives were evaluated for statistical item analysis. The quality of MCQ items was assessed by difficulty index (DIF), discrimination index (DI), and distractor efficiency (DE) using students' exam responses. Microsoft excel sheet and SPSS version 26 were used for data management and analysis.
Results: Post-exam item analysis showed that 11 (22%) and 22 (44%) MCQs had too difficult and poor ranges for difficulty and discriminating powers respectively. The overall DE was 71.3%. About forty (20%) distractors were non-functional. Only 8 (16%) MCQs fulfilled the recommended criteria for all-DIF, DI, and DE parameters.
Conclusions: The desirable criteria for quality parameters of MCQ items were satisfied only in a few items. The result implies the need for quality improvement. Continuous trainings are required to improve the instructors' skills to construct quality educational assessment tools.

Keywords: Multiple choice question, DIF, DI, DE, Ethiopia

INTRODUCTION

The teaching-learning process requires continuous monitoring and evaluation of learning activities.[1] Testing students is essential to enhance communication, motivate them to study hard, identify remediation needs, make promotion decisions, and identify curriculum-related pitfalls.[2] Different types of exam formats are used for educational assessments. Multiple choice questions (MCQs) are health education's most frequently utilized assessment tools. It is a golden standard tool for usual classroom tests to national exams.[1,3,4] MCQ item is preferably used to assess cognitive, affective, and psychomotor domains of learning as it is objective, comparable, and has reduced tendency of examiners' bias.[5] A well-constructed MCQ items are preferred to produce the most reliable scores among students because they lead less to guessing answers than true-false exam items.[6]

An MCQ item has two components: a question or an incomplete statement called a stem and a list of choices called options. A list of options has a single correct choice (key) and several incorrect alternatives called distractors.[4,6,7] Constructing high-quality distractors is an essential aspect of developing MCQ items for testing

students. Ideally, the number of options in MCQ items should be as many as possible. However, nowadays it has been recommended that the MCQs item comprises three to four options to the minimize non-functioning distractor.[8]

MCQ items must be developed by considering their reliability and validity. There are different strategies for assuring the quality of MCQ items regarding reliability and validity. The most widely applied and effective method is an item-analysis process that checks the reliability and validity of MCQ items by measuring three parameters: DIF, DI, and DE. The DIF provides information about the easiness or difficulty of a question asked to students to answer. The DI provides information to distinguish the students having high scores about the subject matter assessed from those having low scores. The third parameter, DE, tells the plausibility of distractors. It is particularly used to count the proportions of non-functional distractors (NFD).[8,9] The validity of educational assessment could be maintained by involving MCQ items having a mid-range difficulty and high discriminating indices with functional distractors.[10]

Brief, clear, and discriminating MCQ items must be constructed for the effective evaluation of health science students.[4,11] Appropriately constructed MCQ items are required to avoid confusion among examinees and to make them score optimally to educational objectives.[6] If most MCQ items are answered correctly by 30-80% of students, then the test could be considered good in its quality.[12] However, constructing appropriate MCQ items is not an easy task, it is a challenging and time-consuming part of the teaching-learning process.[1] It requires scientific creativity.[4] Educators often face the difficulty of developing credible distractors in MCQ items.[8] Poorly constructed MCQ items make an exam easier or too difficult to be answered correctly by students as intended learning objectives.[9] The validity of the educational assessment is attributed to the content of the exam, quality of exam items, number of exam items, nature of item writing flaws, and psychometric characteristics of items.[13-16]

Even though assessing DIF, DI, and DE using item analysis is pivotal to checking the quality of MCQ items, it is often overlooked by many educators.[10] In response to the nation's call for quality education, Injibara university developed an integrated and competency-based curriculum. A variety of assessment methods are used in the examination. However, little is known about the qualities of MCQ items of the exam. Therefore, this study aimed to assess the quality of the MCQ items in response to three parameters (DIF, DI, and DE) prepared by public health instructors at the college of medicine and health sciences, Injibara University, northwest Ethiopia. The study was intended to answer the question "Are the psychometric qualities of the MCQ items corresponding with the desired values for the DIF, DI, and DE?

METHODS

A descriptive cross-sectional study was conducted from May 1-15, 2023 among the MCQ items of public health courses that were administered to the 2nd year nursing students at Injibara university, Ethiopia. All forms of exam items administered to 2nd year nursing students at Injibara university in the first mid-term of the 2023 academic year with the respective responses of examinees were considered. The tests of three public health courses were randomly selected using a lottery method. The tests comprised 50 MCQ items with 200 alternatives (50 keys and 150 distractors). The data were collected from the selected MCQ test items using a Microsoft excel spreadsheet. Data were exported to SPSS version 26 for item analysis. Figures and tables were used to present the results. The MCQ items were evaluated through item analysis for three parameters (DIF, DI, and DE) using the item response theory (IRT) model.[17]

Operational definition

DIF: Measures the easiness or difficulty of question items. It was calculated after grouping one-third of students as high scorer group (H) and the other one-third as low scorer group (L) using the formula ((H+L)/T)×100% where H is the number of students who got the correct answer among high scorers, L is the number of students who got the correct answer among low scorers. T is the total number of students in both groups including non-responders. Then, the indices value of >70%, 30-70%, and <30% findings were considered as too easy, desirable, and too difficult levels of question items respectively.[9]

DI: Measures the tendency of question items to distinguish the high scorer students from the low scorer ones. It was calculated using the formula 2×((H-L)/T). Then, ≤0.2, 0.21-0.24, 0.25-0.35, and ≥0.36 findings were considered as the poor, acceptable, good, and excellent status of question items respectively.[9]

DE: Indicates whether or not the MCQ items contain NFD. A distractor that was chosen by less than 5% of students was considered non-functional. If an MCQ item has 3 or more NFD, its efficiency is zero, and the presence of 2, 1, and 0 NFD implies 33.3%, 66.6%, and 100% DE respectively. The range is always from 0% to 100%.[9]

RESULTS

A total of 50 MCQ items having 200 options were incorporated into our study. The quality of each item was assessed in response to item analysis. A total of 38 2nd year nursing students' response to the tests of three public health courses was used to review the quality of MCQ items for the item response theory model. The proportion of students who got the correct responses to the total MCQ items accounted for 62.9% (SD=21.14).

DIF

The result showed that 39 (78%) MCQ items had difficulty power in the desirable range (30-70%). None of the items was found to be in the too-easy (>70%) range of power. However, 11 (22%) MCQ items were found to be too difficult (having the power of less than 30%) to be answered by the students. The overall mean of DIF was 41.37% (SD=14.55) (Table 1).

DI

The majority of 44% proportion of MCQ items were found to be poor in their power of discriminating against students. Only 6 (12%) and 7 (14%) items had discrimination power of excellent and acceptable ranges respectively. 2 (4%) MCQ items showed negative discriminating power. The overall mean of discriminating power accounted for 0.204 (SD=0.13) (Table 1).

DE

Out of 50 items 20 (40%) had been constructed with effective distractors. However, 7 (14%) and 3 (6%) MCQ items were revealed as having 33.3% and 0% DE respectively. These reflected that 20% of the total MCQ items had poor levels of DE. The overall items had 71.3% (SD=29.4) levels of distractor effectiveness. Regarding to alternatives in MCQ items, about 40 (20%) alternatives were chosen by less than 5% of students making them NFD (Table 1).

In combination with all parameters, overall quality assessment regarding DIF, DI, and DE indicated that only 8 (16%) MCQ items fulfilled satisfaction to the recommended criteria of these parameters (Figure 1).

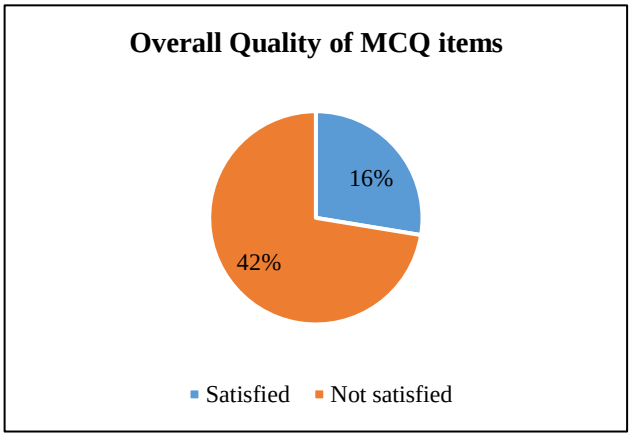

Figure 1: Proportion of MCQ items satisfying the desirable criteria for all DIF, DI, and DE parameters.

Table 1: DIF, DI, and DE of MCQ items (n=50).

Parameters	Ranges	Interpretation	N	Percentages (%)
DIF	>70%	Too difficult	0	0.0
	30-70%	Desirable	39	78.0
	<30%	Too easy	11	22.0
DI	<0.20	Poor	22	44.0
	0.21-0.24	Acceptable	7	14.0
	0.25-0.35	Good	15	30.0
	>0.36	Excellent	6	12.0
DE	0%	3 or more NFD	3	6.0
	33.3%	2 NFD	7	14.0
	66.6%	1 NFD	20	40.0
	100%	0 NFD	20	40.0

Table 2: Comparison of DIF, DI, and DE of MCQ items.

Parameters	DIF	DI	DE
Range	5.26-65.79	-0.16-0.53	0-100%
Mean (SD)	41.37% (14.55)	0.204 (0.13)	71.3% (29.4)

DISCUSSION

Our study evaluated exam items to measure whether the MCQ items rely on the recommended ranges for DIF, DI, and DE parameters. Regarding the DIF, the present study revealed that about 39 (78%) MCQ items had an acceptable range of DIF. This finding is approximately in line with other findings in Islamabad and Bangalore which reported 53 (81%) and the 31 (62%) MCQ items relying upon the acceptable range of DIF respectively.[9,10] However, it was higher than the finding of a study conducted by Patel which evaluated 40 MCQ items and reported that only 18 (45%) items relied upon the acceptable range.[18] Our study showed that about 11 (22%) MCQ items had a DIF of less than 30% reflecting the status of questions being too difficult. This is significantly higher than other findings.[1,18] On the other hand, no MCQ items were relied on in the too-easy range of the DIF which is inconsistent with other findings.[1,5,18] The overall mean DIF in the present study was 41.37% (SD=14.55). This is somewhat lesser than the findings in Karamsad and India from which the means of 55.9%

(SD=15.7) and 55.32% (SD=7.4) were reported for the DIF, respectively.[18,19]

In the present study, the proportion of MCQs with poor power of discrimination accounted for 44% (22 MCQs). This is lower than a finding in Pakistan from which the power of discrimination in the poor range accounted for 67.5%.[20] However, it is higher than the findings of two studies in India in which only 20% and 14% of MCQs had poor discriminating power, respectively.[19,21] Our study reflected that only 7 (14%) MCQs had an acceptable range of discriminating power which is nearly congruent with another finding.[9] Overall, 2 MCQ items showed negative discriminating power and it is consistent with another finding.[22] The overall mean of discriminating power was 0.204 (SD=0.13) which is almost consistent with another finding in India, but lower than other findings.[9,19,22] The proportion of students who got the correct responses to the total MCQ items accounted for 62.9% (SD=21.14). This implies that the MCQ items were averagely good in response to discriminating examinees corresponding with the recommended guideline.[12] Regarding distractors, our study revealed that 20 (40%) MCQs had effective distractors (DE=100%). It nearly corresponds with the result of another study by Kolte in which about 47.5% of MCQs had effective distractors.[23] A somewhat inconsistent result was reported in Islamabad from which the proportion of MCQs with effective distractors accounted for 25%.[9] Out of 200 total distractors, about 160(80%) were functional. This finding is quite consistent with the other result reported in India from which the functional distractors accounted for 82%.[21] The overall mean for DE was 71.3% (SD=29.4). It nearly corresponds with other findings by Gajjar et al and Kolte from which 88.6% and 75.55% of DE were reported.[5,23] However, it is significantly higher than the result of another study in India which revealed a mean of 32.35% for DE.[19]

To satisfy the desirable (acceptable) level, the MCQ items should have a DIF of 30-70% range with a DI exceeding 0.21 and a DE of 100%.[9,24] In our study, only 8 (16%) MCQs satisfied the fulfillment for these three criteria. This result is comparable to the result of the study in Islamabad in which the proportion of MCQs satisfying the three criteria accounted for 15%.[9] The finding implies the need for quality improvement in developing MCQ items.

CONCLUSION

The fulfillment of the desirable value for all parameters of quality for MCQs was satisfied only in a few items implying the need for quality improvement. Continuous training and workshops regarding the quality of exam items are imperative to develop the skills of instructors to construct quality exam items. The college of medicine and health sciences is recommended to design strategies and activities concerning this issue to improve the quality of educational assessment.

ACKNOWLEDGMENTS

Authors would like to thank to instructors of public health courses for their willingness to give the students' test papers.

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

REFERENCES

1. Mehta G, Mokhasi V. Item analysis of multiple choice questions-an assessment of the assessment tool. Int J Health Sci Res, 2014;4(7):197-202.
2. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences, 3rd edittion, National Board of Medical Examiners Philadelphia. 1998.
3. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educator. 2012;37(3):98-104.
4. Palmer E, Devitt P. Constructing multiple choice questions as a method for learning. Ann Academy Med Singapore. 2006;35(9):604.
5. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. Indian J Community Med. 2014;39(1):17-20.
6. Burton SJ, Sudweeks RR, Merrill PF, Wood B. How to prepare better multiple-choice test items: Guidelines for university faculty. Brigham Young University, Department of Instructional Science. 1990.
7. Coughlin P, Featherstone C. How to write a high quality multiple choice question (MCQ): a guide for clinicians. Eur J Vascular Endovascular Surg. 2017;54(5):654-8.
8. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Med Educat. 2009;9(1):1-8.
9. Mahjabeen W, Alam S, Usman H, Tahira Z, Rubab B, Sadaf K, et al. Difficulty index, discrimination index and distractor efficiency in multiple choice questions. Ann PIMS-Shaheed Zulfiqar Ali Bhutto Med University. 2017;13(4):310-5.
10. Christian DS, Prajapati AC, Rana BM, Dave VR. Evaluation of multiple choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat. Int J Community Med Public Health. 2017;4(6):1876-81.
11. Brady AM. Assessment of learning with multiple-choice questions. Nurse Educat Pract. 2005;5(4):238-42.

12. Kehoe J. Basic item analysis for multiple-choice tests. Practical Assess Res Evaluat. 1994;4(1):10.
13. Ali SH, Carr PA, Ruit KG. Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. J Scholarship Teaching Learning. 2016;16(1):1-14.
14. Biggs J, Tang C. E-book: Teaching for Quality Learning at University. McGraw-hill education (UK). 2011.
15. Epstein RM. Assessment in medical education. N Eng J Med. 2007;356(4):387-96.
16. Tavakol M, Dennick R. The foundations of measurement and assessment in medical education. Med Teacher. 2017;39(10):1010-5.
17. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: an illustrative guide: AMEE guide no. 72. Med Teacher. 2013;35(1):e838-48.
18. Patel RM. Use of Item analysis to improve quality of Multiple Choice Questions in II MBBS. J Educat Technol Heal Sci. 2017;4(1):22-9.
19. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivasd V. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Med J Armed Forces India. 2021;77(1):S85-9.
20. Rehman A, Aslam A, Hassan SH. Item analysis of multiple choice questions. Pak Oral Dental J. 2018;38(2):291-3.
21. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. Int J Appl Basic Med Res. 2016;6(3):170.
22. Shete AN, Kausar A, Lakhkar K, Khan ST. Item analysis: An evaluation of multiple choice questions in Physiology examination. J Contemporary Med Education. 2015;3(3):106-9.
23. Kolte, V., Item analysis of multiple choice questions in physiology examination. Indian J Basic Appl Med Res. 2015;4(4):320-6.
24. Namdeo SK, Sahoo B. Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India. Int J Res Med Sci. 2016;4(5):1716-9.

Cite this article as: Alemu AT, Tesfa H, Mulugeta A, Fenta ET, Belay MA. Quality of multiple-choice question items: item analysis. Int J Sci Rep 2024;10(6):195-9.