ANALYSIS OF ONE-BEST TYPE MULTIPLE CHOICE QUESTIONS AT ARMY MEDICAL COLLEGE, RAWALPINDI

Mubashir Sharif, Yawar Hayat Khan*, Rahila Yasmin*, Amna Amjad, Ayesha Aslam

Armed Forces Institute of Dentistry/National University of Medical Sciences (NUMS) Rawalpindi Pakistan, *Islamic International Medical & Dental College, Islamabad Pakistan

ABSTRACT

Objective: To assess the quality of multiple choice questions used for evaluation of 2nd year dental students' knowledge regarding pre-clinical prosthodontics.

Study Design: Cross sectional study.

Place and Duration of Study: Army Medical College/Armed Forces Institute of Dentistry, Rawalpindi, form April 2019.

Methodology: The study included 48 undergraduate students of 2nd year BDS at Army Medical College/Armed Forces Institute of Dentistry in April 2019, comprising of 25 one – best type multiple choice questions (MCQs), each having four options a-d. Parameters employed in the item analysis included difficulty index, discrimination index (DI) and the distractor efficiency (DE). Data was analyzed using SPSS version 24.

Results: Difficulty index showed that 20% multiple choice questions were "too difficult". Fifty two percent of multiple choice questions had a poor discrimination index and of the 100 distractors, 29 (29%) were non-functional. Items with moderate difficulty, higher discrimination and functional distractors must be incorporated in multiple choice questions to improve the test standard and quality.

Conclusion: Items administered in this test were of moderate quality. Flawed items require careful revision to improve their quality before using them in any future assessments.

Keywords: Analysis, Discrimination index, Educational assessment, Examination questions, Functional distractors.

INTRODUCTION

Assessment is an integral part of student learning. It involves testing, measuring, collecting, combining information, and providing feedback. Assessment tools have to be valid, objective and reliable[1]. A number of different assessment tools are available to assess students' learning and performance. These include multiple choice questions, extended matching questions, true/false[2], short answer questions, short essay questions, long essay questions, modified essay questions[3]. Choice of an assessment method is based on the specific learning objectives to be assessed as well as the reliability, validity and feasibility of the assessment tool[4].

MCQs not only aim to assess factual knowledge, but also measure other objectives within Bloom's taxonomy of learning, such as comprehension, application, analysis, synthesis and evaluation[5]. Although, It is a laborious task to make a high quality MCQ based examination yet it is comparatively a better assessment tool than other methods as it leaves little room for bias[6]. Also the scoring of an MCQ examination is objective and more reliable[7]. Most of examinations these days comprise of type A MCQs consisting of a problem–statement followed by four or five options[8].

Item analysis provide information regarding the reliability and validity of a test item. The decision to keep, review or discard an item from the test is based on the statistical analysis of students' performance in the examination[9]. Common parameters employed in the item analysis include difficulty index (DIFI), discrimination index (DI) and the distractor efficiency (DE). DIFI ranges between 0 and 1 (0%-100%) while the recommen-

Correspondence: Dr Amna Amjad, Department of Prosthodontics, AFID, Rawalpindi Pakistan

Received: 25 Feb 2020; revised received: 09 May 2020; accepted: 12 May 2020

ded DIFI for an item in the MCQ is between 30-70% which means items of moderate difficulty. DI ranges between 0 and 1 and an acceptable DI for an item should be more than 0.2. DE of a well-constructed item should be 100%. Any distractor that has been selected by <5% of students is considered a non-functional distractor (NFD). NFDs should be reviewed, revised and replaced[10].

The objective of this study was to assess the quality of one best type MCQs used to evaluate 2nd year BDS students in the subject of Pre-clinical Prosthodontics by conducting a post – test item analysis and calculating difficulty Index, discrimination index and distractor efficiency. The outcome of the study will help us better understand and improve the quality of student assessment tool making it more reliable and acceptable.

METHODLOGY

This was a cross sectional study carried out at Army Medical College/Armed Forces Institute of Dentistry, Rawalpindi in April 2019 after approval from the Institutional Ethics Review Board (IERB: 905/Trg-ABP1K2). The study was carried out on 48 undergraduate students of 2nd year BDS, this sample was selected as it was the strength of the entire batch. Informed consent was taken from these students for inclusion into this study. They were asked to complete an assessment comprising of 25 one-best type MCQs. Each question had a stem and four responses. Each correct answer was given a score of 1 while there was no negative marking. Post–validation of the test was done by item analysis. Difficulty index was calculated by the formula

$$P = (H+L/N) \times 100$$

Where H is the number of students in higher ability group correctly answering the item and L is the number of students in lower ability group correctly answering the item and N is the total number of students. Discrimination Index was calculated by the formula

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

Distractor efficiency, number of non-functional distractors (NFD) per item and number of items with non-function distractors were also calculated. Data was analyzed using SPSS version 24. Mean ± standard deviation for all three parameters (DIFI, DI and DE) were calculated. Percentage of items falling in various categories of difficulty and discrimination were also calculated.

RESULTS

Forty eight students took the test which comprised of 25 MCQs. Table-I highlighted the mean of the three parameters of item analysis. Results of DIFI showed that 20% MCQs were "too difficult" with DIFI less than 30% (fig-1). 52% of MCQs had a poor discrimination index with DI <0.2 (fig-2). Of the 100 distractors, 29 (29%) were non-functional (table-II). Out of total 25 MCQs, only 6 (24%) had no non-functional distractors (table-III).

Table-I: Mean values of various indices used in item analysis.

Parameter	Mean ± SD
Difficulty Index	37.23 ± 11.62
Discrimination Index	0.17 ± 0.11
Distractor Efficiency	61.33 ± 29.94

SD: Standard Deviation

Table-II: Frequency of non-functional distractors in test items.

Distractor Analysis	
Total no. of items	25
Total no. of distractors	100
Functional Distractors	71 (71%)
Non-Functional Distractors	29 (29%)

Table-III: Distribution of test items according to the frequency of non-functional distractors.

No. of Non – Functional Distractors	n (%)
No NFD 0	6 (24%)
1 NFD	11 (44%)
2 NFD	6 (24%)
3 NFD	02 (8%)

DISCUSSION

A well- structured assessment helps evaluate how well learning outcomes have been achieved, allows the teacher to test higher levels

of cognitive domain and aids in distinguishing between high and low achievers[11]. The assessment tool of any examination should be designed according to the objective[12].

If properly designed, one-best MCQs are one of the most valid assessment tools that quickly assess any level of cognition according to Bloom's taxonomy[13]. Post-test analysis of the assessment

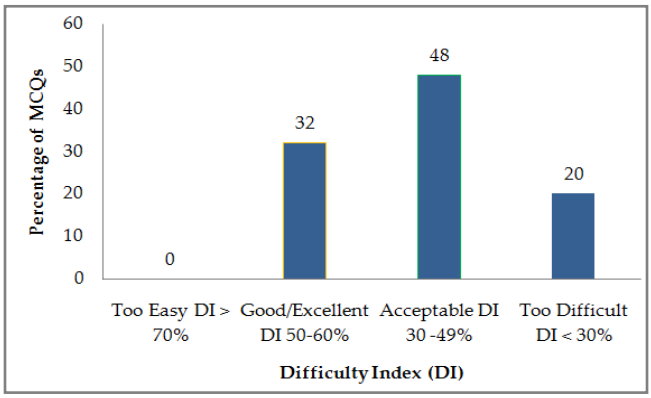

Figure-1: Percentage distribution of items (MCQs) according to their difficulty index.

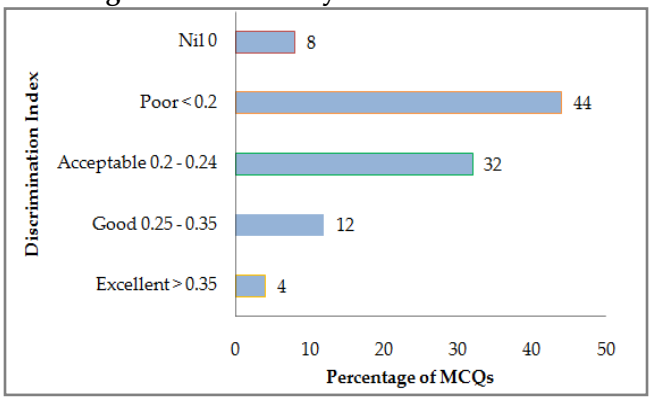

Figure-2: Percentage distribution of test items according to their discrimination index.

aids in reviewing the quality and performance of test items and helps improve the items for future use. Acceptable range for DIFI is 30-70%. The mean DIFI in the present study was 37.23 ± 11.62 indicating that on average, items were of "acceptable difficulty" to attempt for the students. About 80% of the items in the present study were in the acceptable range, while 20% were in the "too difficult" range.

Difficult items must be thoroughly reviewed for any ambiguity, controversial alternatives, or even an incorrect key. Hingoro *et al* reported comparable results with 78% of its items in the acceptable range, consistent to the results of this study[14]. On the contrary, Shete *et al* reported 30% of their items in the acceptable range, consistent to the results of this study[15] while Rehman *et al* reported 52.5% of test items as "too difficult". Mean DI was 0.17 ± 0.11. 48% of the items had a DI >0.2, with 4% showing excellent discrimination (DI>0.35) while 44% of the test items had a DI <0.2 (poor discrimination). 8% of the test items failed to discriminate between high achieving and low achieving students (DI=0). None of the items had a negative DI[16].

Namdeo *et al* reported a mean DI 0f 0.33 ± 0.23 which signifies good discriminative ability of test items[17]. Hingorjo *et al* reported a mean DI of 0.36 ± 0.17 with only 2 out of 50 items showing a negative DI[14]. Rehman *et al* reported a mean DI of 0.12 ± 0.13 with 67.5% items having poor discrimination (DI <0.2)[16]. A negative DI may be attributed to an incorrect key, ambiguous structure of an item and poor preparation of students. No discrimination may result if the item is too easy if it is too difficult or flawed[18].

Mean distractor efficiency (DE) in the study was 61.33 ± 29.94%. Of the 100 distractors, 29% were NFDs. Only 24% items had no NFDs while 44% of items had at least 01 NFD. Presence of non-functional distractors increases the DIFI, making an item easier to attempt. Subsequently, the DI of an item with greater number of NFDs will be poor.

ACKNOWLEDGEMENT

We would like to acknowledge all the study participants for their cooperation.

RECOMMENDATIONS

The results calculated manually should be compared with those calculated using Optical mark recognition (OMR) software and finalization of results after excluding the flayed item and re administration of test items is recommended.

CONCLUSION

It can be concluded that items administered in this test were of moderate quality. Flawed

items requires careful revision to improve their quality before using them in any future assessments. Items with moderate difficulty, higher discrimination and functional distractors must be incorporated in assessments to improve the assessment standard and quality.

CONFLICT OF INTEREST

This study has no conflict of interest to be declared by any author.

REFERENCES

1. Case SM, Swanson DB. Extended-Matching Items: A practical alternative to free-response questions. Teaching Learn Med 1993; 5(2): 107–15.
2. 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): 1–8.
3. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. Int J App Basic Med Res 2016; 6(3): 170.
4. 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 Commu Med 2014; 39(1): 17-23.
5. Skakun EN, Nanson EM, Kling S, Taylor WC. A preliminary investigation of three types of multiple choice questions. Med Edu 1979; 13(2): 91–96.
6. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos Uni Med J 2018; 18(1): e68-74.
7. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Edu 2017; 4(1): 1301013.
8. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in Physiology. South East Asian J Med Educ 2013; 7(1): 45-50.
9. Mehta G, Mokhasi V. Item analysis of multiple choice questions-an assessment of the assessment tool. Historical Aspects Leech Ther 2014; 4(7): 1–3.
10. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Edu Measure Issues Prac 2005; 24(2): 3–13.
11. Sayyah M, Vakili Z, Masoudi Alavi N, Bigdeli M, Soleymani A, Assarian M, Azarbad Z. An item analysis of written multiple-choice questions: Kashan Uni Med Sci. Nurs Midwifery Stud 2012; 1(2): 83-87.
12. Nwadinigwe PI, Naibi L. The number of options in a multiple-choice test item and the psychometric characteristics. J Educ Pract 2013; 4(1): 189-96.
13. Armstrong P. Bloom's taxonomy. Vanderbilt University Center for Teaching. 2016.
14. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc 2012; 62(2): 142–47.
15. Shete A, Kausar A, Lakhkar K, Khan S. Item analysis: An evaluation of multiple choice questions in Physiology examination. J Contemp Med Edu 2015; 3(3): 106-12.
16. Rehman A, Aslam A, Hassan SH. Item analysis of multiple choice questions. Pak Oral Dental J 2018; 38(2): 291–93.
17. 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): 1716-19.
18. 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.