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, from 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. A number of different assessment tools are available to assess students’ learning and performance. These include multiple choice questions, extended matching questions, true/false, short answer questions, short essay questions, long essay questions, modified essay questions. 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.

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. 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. Also the scoring of an MCQ examination is objective and more reliable. Most of examinations these days comprise of type A MCQs consisting of a problem-statement followed by four or five options.

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. 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-
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.

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

\[ P = \frac{(H+L)}{N} \times 100 \]

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

\[ DI = \frac{(H-L)}{N} \times 2 \]

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).

**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\textsuperscript{11}. The assessment tool of any examination should be designed according to the objective\textsuperscript{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\textsuperscript{13}. Post-test analysis of the assessment 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\textsuperscript{14}. On the contrary, Shete et al reported 30% of their items in the acceptable range, consistent to the results of this study\textsuperscript{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\textsuperscript{16}.

Namdeo et al reported a mean DI of 0.33 ± 0.23 which signifies good discriminative ability of test items\textsuperscript{17}. Hingorjo et al reported a mean DI of 0.36 ± 0.17 with only 2 out of 50 items showing a negative DI\textsuperscript{14}. Rehman et al reported a mean DI of 0.12 ± 0.13 with 67.5% items having poor discrimination (DI <0.2)\textsuperscript{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\textsuperscript{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**