Article published online: 2020-04-22

NUJHS Vol. 5, No.4, December 2015, ISSN 2249-7110

Nitte University Journal of Health Science

Original Article

Role of item analysis in post validation of multiple choice questions in formative assessment of medical students

Sajitha K.[1], Harish S. Permi[2], Chandrika Rao[3] & Kishan Prasad H.L.[4]

[1,3]Assistant Professors, [2,4]Associate Professors, Department of Pathology,
K.S. Hegde Medical Academy, Nitte University, Mangalore, Karnataka, India.

Correspondence
Sajitha K.
Assistant Professor, Department of Pathology, K.S. Hegde Medical Academy, Nitte University, Mangalore, Karnataka, India.
Mobile : +91 97419 93622 E-mail : drsk29@hotmail.com

Abstract

Background: Multiple choice questions (MCQ) are used in the assessment of students in various fields. By this method of assessment it is possible to cover a wide range of topics in less amount of time. However the reliability of the test depends on the quality of the MCQ. The MCQ can be evaluated based on the Difficulty Index (DIF I), Discriminatory Index (DI) and Distracter Efficiency (DE).

Objectives : To evaluate the MCQs based on the Difficulty Index (DIF I), Discriminatory Index (DI) & Distracter Efficiency (DE) and develop a valid pool of questions.

Also to assess learner performance and discriminate between students of higher and lower abilities.

Materials and Methods: A total of 120 students were assessed based on multiple choice questions in pathology. The number of items were 20 and the number of distracters were 60. Data was entered and analyzed in MS Excel 2007 and simple proportions, mean and standard deviations were calculated.

Results: Mean and standard deviations for DIF I, DI and DE were 57.8 ± 17.4%, 0.27 ± 0.17 and 84.98 ± 20.2 % respectively. Out of the 20 items, 11 items had good level of DIF I (31 – 60%), eight (8) items were considered easy (DIF I ⩾ 61%) and one (1) item was considered difficult (DIF I ⩽ 30). Mean DI in present study was 0.27 ± 0.17. Analysis of the DI showed good discrimination power in eighteen (18) of the items. Out of the 60 distracters, nine (9) were non – functional distracters (NFD) and were seen in eight items. Out of these, seven items had one NFD each and one item had two NFD.

Conclusions: The study emphasizes on the importance of use of item analysis in construction of good quality MCQs and also in the evaluation of learner performance.

Keywords: Item, Discrimination Index, Difficulty Index, Distracter

Introduction

According to the recent concept of Competency Based Medical Education (CBME) given by the Medical Council of India, it is an outcome based model which measures competency, which are guided by the needs of the community and until achieved by the learners[1]. In this regard, student assessment should be done by using a measurable standard method. Regular examination of student performance forms a part of the instruction process and helps to alter and revise the methods of teaching as well as learning.

Access this article online
Quick Response Code

Multiple choice questions (MCQ) have been used in the assessment of students in the medical field since 1999 in both departmental as well as university examinations.[2] By this method of assessment it is possible to cover a wide range of topics in less amount of time. A well constructed MCQ tests the high cognitive level processes rather than recollection of memorized facts. Designing appropriate MCQ to correctly evaluate the level of knowledge of the students is a laborious task. And the reliability of the test depends on the quality of the MCQ.

One of the tools used in evaluation of a test process is an item analysis. The item analysis provides information about the reliability and validity of test items and learner

performance. It serves two purposes.[1] It helps to identify defective test items and secondly, to precisely find the learning materials that the students have and have not mastered, particularly what skills they lack and what material still causes them difficulty. The MCQ can be evaluated based on the Difficulty Index (DIF I), Discriminatory Index (DI) and Distracter Efficiency (DE).[3]

Hence, this study was conducted in our department of pathology to assess and validate the MCQs for second year students. And for assessment of students without being influenced by the performance of other students.

Objectives

1. To evaluate the MCQs based on the Difficulty Index (DIFI), Discriminatory Index (DI) & Distracter Efficiency (DE) and develop a valid pool of questions.
2. Also to assess learner performance and discriminate between students of higher and lower abilities.

Methodology

A total of 120 students of 2nd MBBS were evaluated by MCQs in pathology. It formed a part of a formative assessment which consisted of a 3 hour written paper with MCQs to be completed in the first 20 minutes. The MCQs were pre – validated by peer review. The number of items were twenty (20) and the number of distracters were sixty (60). Each correct response was awarded one mark and the incorrect responses were awarded zero points. The selection of the upper and lower groups were based on Kelley's derivation.[4] The forty (40) learners with the highest test scores were included in the upper criterion group and the forty (40) learners with the lowest test scores in the lower criterion group. The middle group (40) were set aside.

The Difficulty Index (DIF I), Discriminatory Index (DI) & Distracter Efficiency (DE) of each item were calculated.

DIF I describes the percentage of students who answered the item correctly and ranges between 0 and 100%.

DI is the ability of an item to differentiate between students of higher and lower abilities and ranges between 0 and 1. The higher the value of DI, the greater the ability of the item to discriminate between students of higher and lower learning abilities.

These were calculated by the following formulae.[3]

1. **DIF I or p value = [(H + L) / N] × 100**
2. **DI = 2 × [(H – L) / N]**

N = total number of students in both high and low groups and H and L are the number of correct responses in high and low groups, respectively.

Interpretation of the data was done as follows [3] –

Percentage Range(DIF I)	Interpretation	Discrimination Index(DI)	Interpretation
⩾61	Easy	⩾0.25	Excellent
31 – 60	Good	0.15 – 0.24	Good
⩽ 30	Difficult	< 0.15	Poor

An item contains a stem and four options including one correct (key) and three incorrect (distracter) alternatives. A non functional distracter (NFD) in an item is option (s) (other than key) selected by <5% of students.[3] The items were analyzed for distracter effectiveness (DE) based on the number of non-functional distracters (NFD).

Data was entered and analyzed in MS Excel 2007 and simple proportions, mean and standard deviations were calculated.

Results

A total of 20 items and 60 distracters were analyzed.

Mean and standard deviations for DIF I, DI and DE were 57.8 ± 17.4%, 0.27 ± 0.17 and 84.98 ± 20.2 % respectively (Table 1).

Out of the 20 items, 11 items had good level of DIF I (31 – 60%) and could be added to the question bank (Table 2). Eight (8) items were considered easy (DIF I ⩾61%) and one (1) item was considered difficult (DIF I ⩽ 30).

Value of DI normally ranges between 0 and 1. Mean DI in present study was 0.27 ± 0.17 (Table 3). Analysis of the DI showed good discrimination power in eighteen (18) of the items. Based on this, the 18 items can be considered ideal for the question bank.

A higher DE indicates that the set of items were difficult. Mean DE in present study was 84.98 ± 20.2%.

Out of the 60 distracters, nine (9) were non – functional distracters (NFD) and were seen in eight items (Table 4). Out of these, seven items had one NFD each and one item had two NFDs. The remaining 12 items had no non – functional distracters.

Among the nine (9) non – functional distracters, one was seen with an item with a good DIF I while the remaining 8 were seen with the easy items (Table 5). The presence of one or more NFD in an item increases DIF I and makes the item easy. The presence of NFD probably made the items easy in the study. The one item with good DIF I which was considered acceptable for the question bank has to be revised due to the presence of the NFD.

Among the 18 items with a good discrimination index, 8 items showed non – functional distracters and hence had to be revised or discarded. The NFD were equally distributed among the good & excellent items (4 each). (Table 6)

Table 1

PARAMETER	MEAN	STANDARD DEVIATION
Difficulty Index (DI) (%)	57.8	17.84
Discrimination Index (DI)	0.27	0.17
Distracter Efficiency (DE) (%)	84.98	20.2

Table 2

Percentage Range (DIF I)	**Items (N=20)**	**Interpretation**	**Action**
⩾61	08	Easy	Revise / Discard
31 – 60	11	Good	Store
⩽ 30	01	Difficult	Revise / Discard

Table 3

Discrimination Index	**Items (N=20)**	**Interpretation**	**Action**
<0.15	02	poor	Discard/revise
0.15 – 0.24	07	good	Store
⩾25	11	excellent	Store

Table 4 : Distracter Analysis

No of Items	20
Total Distracters	60
Functional Distracters	51
Non Functional Distracters	09
Mean Distracter Efficiency	84.98 ± 20.2%

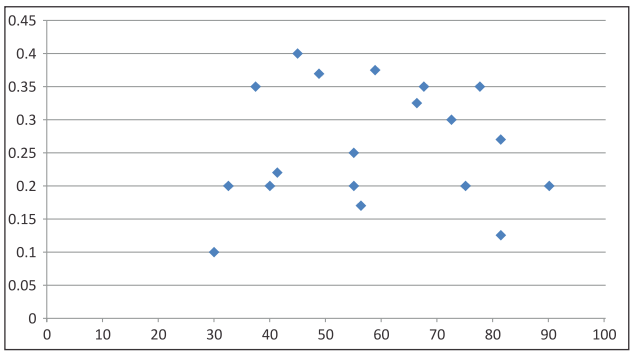
Table 5

DIF I	Items with non functional distracters
⩽ 30	0
31 – 40	0
41 – 60	1
⩾ 61	08

Table 6

Discrimination Index	**Items with NFD**
<0.15	**01**
0.16 – 0.24	**04**
⩾ 0.25	**04**

Fig 1 : Correlation between Difficulty Index and Discrimination Index



Discussion

MCQs are used mostly for comprehensive assessment at the end of a semester or academic sessions and provide feedback to the teachers on their educational actions.[5,6] MCQ based exams form a good method of assessment of knowledge of a subject because of their ability to cover a wide range of topics and due to their objectivity. And MCQs form an important method of evaluation in medical field.

Reliability of a test item is very important in its construction.[6] An item analysis can provide useful information for improving the quality and accuracy of multiple-choice or true / false items. Item difficulty (DIF I) indicates the percentage of students that correctly answered the item and is also referred to as the P-value.[3,6] A

high difficulty index indicates an easy set of questions. However if these questions measure a valid performance standard, they could still be used as good test items. A low difficulty index indicates a difficult item and should be reviewed for the use of confusing language. It should be either removed in subsequent tests or should be identified as an area for re- instruction. It also may indicate that the topic tested is inappropriate at that level for the students.[7,8]

Discrimination Index or Point Biserial is a statistic which indicates the extent to which an item has discriminated between the high scorers and low scorers on the test.[4] An ideal item should have a positive discrimination index of at least 0.2.[4] A high discrimination index will help us differentiate between good learners and poor learners and gives an effective feedback to the teachers.

Distracter evaluation also forms another useful item review technique. Distracters are the incorrect alternatives in a multiple-choice item. Non- functional distracters (NFDs) are item options chosen by < 5% of the students.[3]

Items with a moderate level of DIF I and high discrimination index may still be flawed if there are NFD.

The present study assessed the item analysis outcomes of the MCQ in the first sessional examination for Pathology course for MBBS students. The majority of the items in our study was framed from the topics which are essential to be mastered and was of acceptable difficulty level and also showed good discrimination. In our study, 90% of the items had a DI of ⩾ 0.2. Both the easy and moderately difficult items showed good discriminative ability. One difficult item (DIF I < 30%) and one easy item (DIF I– 81.25 %) showed poor discrimination (Fig 1).

Ho et al,[9] in a similar study reported that too easy or too difficult items showed poor discrimination.

In a study by Ghadam et al,[10] who assessed the item analysis outcomes of MCQs in 4 different semesters over 3 years found that good MCQs and improved teaching method based on the item analysis variables were associated with an increased number of students who passed the exam with a greater mean score.

Conclusion

Due to its high discrimination index, the study helps to identify the poor learners. The very easy and difficult items have to be revised and reconstructed. Some of the items with good DIF I and with good DI were not acceptable due to the presence of non - functional distracters. Hence these have also got to be reviewed, reconstructed and revalidated. A regular analysis of the items should be carried out in this manner after every examination to improve the standard of assessment and develop a valid questions bank. Student feedback and peer review will have a sustained positive impact on the quality of MCQ items. In addition to revision of questions, item analysis should also be followed up with improved teaching methods by identifying the poor learners and areas of learner difficulties. Similar analysis of the MCQs can be conducted for the summative evaluation of university examination.

Reference

1. Frank J R, Mungroo R, Ahmad Y, Wang M, de Rossi S, Horsley T. Toward Competency Based Education in Medicine: A systematic Review of Published Definitions. Med Teacher 2010; 32: 631-7.
2. Pande S S, Pande S R, Parate V R, Nikam A P, Agrekar S H. Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. South-East Asian Journal of Medical Education 2013;7(1):45 – 50.
3. Gajjar S, Sharma P, 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.
4. Kelley TL. The selection of upper and lower groups for validation of test items. J Educ Psychol 1939; 30: 17-24.
5. Brown FG. Principles of educational and psychological testing. 3rd ed. New York: Holt, Rinehart and Winston, 1983.
6. Mitra N K, Nagaraja H S, Ponnudurai G, Judson J P. The level of Difficulty and Discrimination Indices In Type A Multiple Choice Questions Of Pre–clinical Semester I Multidiciplinary Summative Tests. IeJSME 2009;3(1): 2-7.
7. Senanayake MP, Mettananda DSG. Standards medical students set for themselves when preparing for the final MBBS examination. Annals Acad Med 2005; 34: 483-85.
8. Marso RN, Pigge FL. An analysis of teacher made tests and item construction errors. J Contemp Edu Psych 1991; 16: 279-86.
9. Ho T.F, Yip W.C.L, Tay J.S.H. The use of multiple choice questions in medical examination: An evaluation of scoring and analysis of results. Singapore Medical Journal 1981; 22(6):361-67.
10. Talebi G A, Ghaffari R, Eskandarzadeh E, Oskouei A E. Item Analysis an Effective Tool for Assessing Exam Quality, Designing of Appropriate Exam and Determining Weakness in Teaching. Res Dev Med Educ 2013; 2(2): 20-23.