ORIGINAL ARTICLE

Item Analysis of MCQs of a Pharmacology Term Exam in a Private Medical College of Pakistan

MUHAMMAD NAUMAN SHAD[1], AJAZ FATIMA[2], SHAHNAZ FATIMA[3], *SADIA CHIRAGH[4]*

ABSTRACT

Background: Multiple choice questions (MCQs) are widely used to assess students in local and university examinations in the medical field. The quality of the MCQs given in these examinations should be assessed from time to time to ensure medical students are given right choice of questions.
Aim: To evaluate quality of MCQs with assessment tools and to store, revise or discard MCQs based on the results.
Methods: The study was conducted in the department of Pharmacology in a private medical college. Thirty students of 3rd Year MBBS appeared in the term exam comprising of 65 MCQs and 10 SEQs, each MCQ having 5 options. Evaluation of the MCQs (items) was done by item analysis. The MCQs were analyzed for difficulty index (p-value), discrimination index (DI), and distractor efficiency (DE). Items having p-value between 30-70 and DI > 0.20 were considered as having good difficulty and discrimination indices respectively. Functional distractors were considered as the ones selected by at least 5% of the students. Items with 100 % distractor efficiency were those with no non-functional distractors (NFDs).
Results: The mean score was 29.5 (out of 65). Difficulty index of 44 items (67.7%) was within the recommended range. 42(64 %) of the items had discrimination index above the recommended value. 31 items (47.7%) showed a good difficulty index as well as a good discrimination index. 45 items (69.2%) had 100% distractor efficiency.
Conclusion: If quality MCQs are given even to students in a private medical college, acceptable values of various item analysis parameters can be produced.
Keywords: Difficulty index, Discrimination index, distractor efficiency, Non-functional distractors

INTRODUCTION

One best multiple choice questions (MCQs) or items are widely being used now a days for assessing knowledge and/or grasp of a subject of undergraduate and postgraduate students. Advantages of MCQs include being very convenient for the students, with a large portion of curriculum being assessed in a short period of time. They are an efficient tool in identifying the weaknesses and strengths in students. They also provide guidelines to teachers on their educational protocols[1].

Testing of higher-order cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, instead of simple recall of isolated facts, can be assessed by preparing properly constructed multiple choice questions [2, 3].

Multiple choice questions are also preferred as a method of assessment because they provide objectivity in assessment, comparability in different settings and minimize examiner's bias. All this is possible if the examiner knows how to correctly formulate a multiple choice question, consisting of a stem, a lead question and several options[4].

Item analysis analyzes student responses to each multiple choice question to assess their quality. This allows for examiners to improve/revise items and the test as a whole. Item analysis enables identifying good MCQs based on difficulty index (DIF I), discrimination index (DI) and distractor efficiency (DE) [5].

Difficulty index is an index which expresses the proportion or percentage of students who answered the item/MCQ correctly. It is frequently called the p-value. The larger the percentage getting an item right, the easier the item & the higher the difficulty index. This is why perhaps "difficulty index" should have been named "easiness index".

The item/MCQ difficulty index is one of the most useful, and most frequently reported, item analyses in statistics. Item difficulty can range from 0.0 (none of the students answered the item correctly) to 1.0 (all of the students answered the item correctly). Experts recommend that the average level of difficulty for an MCQ item should be between 30% and 70%[6].

Discrimination index(DI) is the ability of an item to differentiate between students who topped the test and students who scored the least marks in the test. It ranges between 0 and 1. Higher the value of DI, the more the item is able to discriminate between students of higher and lower abilities. DI of 1 is ideal as it refers to an item which perfectly discriminates between students of lower and higher abilities. DI can sometimes be <0 (negative DI) which simply means that the students of lower ability answered the item more correctly than those with higher ability[7].

Each MCQ is followed by four or five options of which only one is the correct option, with the others being distractors. Non-functional distractor (NFD) is/are an option (s) (other than key) selected by less than 5% of students; alternatively functional or effective distractors are those selected by 5 % or more participants[8].

---

[1]Associate Professor of Pharmacology, Sahara Medical College, Narowal.
[2]Associate Professor of Pharmacology, Lahore Medical & Dental College, Lahore.
[3]Assistant Professor of Pharmacology, Sahara Medical College, Narowal.
[4]Retired Professor of Pharmacology, PGMI, Lahore.
Correspondence to Dr. Muhammad Nauman Shad Email: mnauman2002saj@yahoo.com Cell: 0334-5062652

Distractor efficiency is determined for each item on the basis of the number of NFDs in it and ranges from 0 to 100%. If an item has no NFDs, then it has 100% distractor efficiency and if it has maximum possible NFDs, then it has zero distractor efficiency[9].

MATERIAL AND METHODS

This was a cross-sectional study conducted at Sahara Medical College, Pakistan of 3rd Year MBBS class for the session 2017-18. Thirty out of thirty two students appeared in the term exam. Sixty five MCQs were given from the subject of pharmacology at the end of the second semester of 13 weeks from the sections of CVS, GIT, Respiratory System, Blood and Chemotherapy.

The time given for the MCQ paper was 65 minutes, and was followed by a short essay type paper. The MCQs were of one-best type, having a stem and 5 options, with only one of them being correct and the other 4 were 'distractors'. The students were required to select the correct choice. Each correct response was awarded 1 mark. No mark was given for blank response or incorrect answer. There was no negative marking. Thus, the maximum possible score of the test was 65 and the minimum 0.

Item Analysis: This was done manually. The papers with their responses were divided into three groups:

H High group: top 8 students (around top 27%)

L Low group: bottom 8 students (around bottom 27%) and M (middle group: middle 14 students) after arranging their marks in descending order.

Selection of the upper and lower 27% has been demonstrated to be the most efficient fraction[10].

Difficulty Index: Difficulty index or p-value for each MCQ was calculated using the following formula:

H + M + L/N

In other words, it was calculated as the total number of students from all the three groups who answered the MCQ correctly divided by the total number of students who appeared in the paper. It was taken as a percentage.

Interpretation:

Difficulty index less than 30 % was taken as too difficult

Difficulty index more than 70 % was taken as too easy

Difficulty index ranging from 30 % to 70 % was taken as acceptable (recommended)

Discrimination Index (DI): Discrimination index for each MCQ was calculated using the following formula[9]:

H - L/8

H = Number of top students who answered the MCQ correctly

L = Number of bottom students who answered the MCQ correctly

Interpretation:

DI if negative: very poor discrimination

DI if less than 0.2: poor discrimination

DI if more than 0.2: good/acceptable discrimination

DI if more than 0.4: excellent discrimination

Distractor Efficiency: The number of non-functional distractors (NFDs; for which less than 5% of the class of students opted as the answer) was determined for each MCQ. Items with no NFDs had 100% distractor efficiency; items with one, two or three NFDs had 75%, 50% and 25% distractor efficiencies respectively. Interpretation: The higher the distractor efficiency, the better the quality of the MCQ

Statistical analysis: The data was entered in Microsoft word and Excel 2010 and reported as percentage and/or mean plus or minus standard deviation (SD) of n items.

RESULTS

A total of 30 students gave the exam consisting of 65 MCQs. Scores of the students ranged from 9 to 51 with mean score of 29.5±8.86 (maximum 65 marks). 65 MCQs and 260 distractors were analyzed.

Difficulty Indices: Mean and SD of difficulty indices were 47.6 ±19.5(%) (Table 1). Difficulty indices of 44 out of the 65 MCQs (67.7%) were in the acceptable range (30 – 70 %). 9 MCQs (13.8%) were too easy with difficulty indices more than 70 %. 12 MCQs (18.5%) were too difficult with difficulty indices less than 30% (Table 2).

Discrimination Indices: Mean and SD of difficulty indices were 0.33±0.21 (Table 1). 42 MCQs (64.6%) had acceptable/recommended discrimination indices (> 0.20). 20 MCQs (30.8%) had excellent discrimination indices (>0.40). 23 MCQs (35.4%) had poor discrimination indices, with two of them having very poor (negative) discrimination indices (Table 2).

31 (47.7%) MCQs were 'ideal MCQs' having recommended difficulty indices as well as recommended discrimination indices. 5 (7.7%) MCQs were very poor quality with poor difficulty indices and poor discrimination indices.

NFDs and Distractor Efficiency: Total number of distractors was 260 (65 x 4). Out of these 233 (89.6 %) were functional distractors and 27(10.4%) were NFDs (non-functional distractors). 15 (23.1 %) items had only one NFD. 45(69.2%) MCQs had no NFDs. Thus 45 MCQs (69.2%) had 100 % distractor efficiency. Mean and SD of distractor efficiencies were 89.6±18.2(%).

Table 1: Mean and SD values of Difficulty index, Discrimination index and Distractor efficiency

Parameter	Mean	Standard Deviation
Difficulty Index (%)	47.6	19.5
Discrimination Index	0.33	0.21
Distractor Efficiency (%)	89.6	18.2

Table 2:Distribution of items in relation to DIF I and DI and actions proposed

Cut Off Points	n	Interpretation	Action
Difficulty Index (DIF I)			
< 30 %	12	Too Difficult	Revise or Discard
30 – 70 %	44	Good to Excellent	Store/Save
> 70 %	9	Too Easy	Revise or Discard
Discrimination Index (DI)			
< 0.2	23	Poor	Discard or Revise
> 0.2	42	Good	Store/Save
> 0.4	20	Excellent	Store/Save
< 0 (Negative)	2	Very Poor	Discard or Revise

Table 3: Items (N = 15) with nonfunctional distractors (NFDs) and their relationship with DIF I and DI

DIF I (%)	Items with NFDs	DI	Items with NFDs
< 30	2	< 0.2	9
30 – 70	18	> 0.2	8
> 70	0	> 0.4	3

DISCUSSION

Development of skilled, knowledgeable and competent medical personnel is required for producing good quality medical care. One correct response type MCQs are an efficient tool for evaluation[9]. They measure factual recall and if intelligently constructed can test higher orders of cognition like understanding and application, which is very important for a medical graduate[11]. The method of assessment should be regularly evaluated.

A key part in curriculum development is development of an appropriate assessment strategy. To find out how effective MCQs are in assessing the knowledge of students, it is necessary to evaluate them[11]. The difficulty index, discrimination index and distractor efficiency are major tools used to analyze and assess the quality of MCQs.

Mean difficulty index in this study was 47.6±19.5%, which is within the acceptable range of 41-60%[12]. Various other studies have shown means of difficulty index varying from 39 % to 58 %[13]. In our study, 67.7% of the items were in the acceptable range (30-70%), 13.8% were too easy and 18.5 % were too difficult. Other studies have shown 61-80% items within the acceptable range[13].

Factors causing a low difficulty index include content/learning objective of the MCQs was not covered in class or was out of course, MCQ was beyond the scope of the undergraduate students, MCQ was of a higher cognitive level, ambiguous or unclear due to item flaws, students not prepared enough or of low caliber (as in a private medical college) and other factors[14].

Very difficult items (DIF I <30%) will give low scores, while the easy items (DIF I > 70%) will lead to high scores[15]. Very easy items should be placed either at the start of the exam as 'warm-up' questions or discarded altogether. Difficult items should be reviewed for possible item flaws, or even an incorrect key.

Discrimination index(DI) of an item indicates its ability to differentiate between top performers and bottom performers of the exam. Items that are very easy and thus attempted correctly by most students) will have nil to poor DI. Items that are difficult may give a high DI value with top students performing much better than poor students, but they may also give a low DI value if they are too difficult. Such cases may give a negative DI, in which bottom students perform better than top students.

Mean DI in present study was 0.33 ± 0.21, which ismuch higher than the acceptable cut off point of 0.15[7].30.8 % of the items had excellent DI value over 0.4. Reasons for negative or poor DI can be wrong key, various item flaws[16], or generalized poor preparation of students and other factors mentioned above for poor difficulty index. Items with negative DI mostly need to be revised or discarded, as they decrease the validity of the test[6].

31 out of the 65 MCQs (47.7 %) were ideal in which both DIF I and DI were in the recommended range as compared to 64 % in another study[9]. These are the best type of MCQs that should be saved for use in future exams.

Analysis of distractors is also very important. A distractor chosen by less than 5% of students is a non-functional distractor (NFD) and should be either replaced or reviewed as it affects the overall quality of the question. NFDs make the items easier to answer, thereby affecting the actual assessment of the student. For constructing quality MCQs, plausible distractors should be designed so that number of NFDs is reduced[17].

Mean distractor efficiency (DE) in the present study was 89.6±18.2(%),much higher than DE of 81.4% reported elsewhere in a similar type of study[9].

CONCLUSION

Assessment of MCQs by the above indices highlights the importance of assessment tools for the benefit of both the student and the teacher. This study proves that if quality MCQs are given even to students in a private medical college, acceptable values of various item analysis parameters can be produced; and that too in a challenging subject of pharmacology.

The values of difficulty index, discrimination index and distractor efficiency in this study all were in the acceptable range. Most of the MCQs had acceptable difficulty & discrimination indices and were able to discriminate between very good performing students and the poor students.

REFERENCES

1. Tan LT, McAleer JJ; Final FRCR Examination Board. The introductionof single best answer questions as a test of knowledge in the finalexamination for fellowship of the Royal College of Radiologists inClinical Oncology. *ClinOncol*; 2008; 20: 571-6.
2. Carneson J, Delpierre G, Masters K. Designing and managing MCQs:Vol. 62, No. 2, February 2012 146Appendix C: MCQs and Bloom's taxonomy. (Online) 2011 (Cited 2011Feb 2). Available from URL:http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html .
3. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences, National Board of Medical Examiners 3rd ed. (Online) 2010 (Cited 2011 Feb 5). Available from URL: http://www.nbme.org/publications/item-writing-manual.html
4. Collins J. Education techniques for lifelong learning: writing multiple choice questions for continuing medical education activities and self-assessment modules. *Radiographics* 2006; 26: 543-51.
5. Scorepak®: Item analysis. [Last citied on 2013 Apr 13]. Available from: www.washington.edu./oea/score1/htm
6. Matlock-Hetzel S. Presented at annual meeting of the Southwest Educational Research Association, Austin, January; 1997. [Last cited on 2013 Apr 13]. Basic concept in item and test analysis. Available from: www.ericae.net/ft/tamu/espy.htm
7. Singh T, Gupta P, Singh D. Principles of Medical Education. 3rd ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd; 2009. Test and item analysis; pp. 70–77.
8. 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;9:1–8
9. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. *J Pak Med Assoc*. 2012;62:142–7.
10. Kelley TL. The selection of upper and lower groups for the validationof test items. *J EducPsychol* 1939; 30: 17-24.
11. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQ'sin formative exam in physiology. *South East Asian J Med Educ*2013;7:45-50.
12. Guilbert JJ. 1st ed. Geneva: World Health Organization; 1981. Educational Hand Book for health professionals. WHO offset Publication 35.
13. Rao C, Kishan Prasad HL, Sajitha K,Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *Int J Educ Psychol* Res 2016;2:201-4.
14. Al-Zoubi SM and BaniYounes MA. Low Academic Achievement: Causes and Results. Theory and Practice in Language Studies, Vol. 5, No. 11, pp. 2262-2268, November 2015 DOI: http://dx.doi.org/10.17507/tpls.0511.09
15. Eaves S, Erford B. The Gale group. The purpose of item analysis, item difficulty, discrimination index, characteristic curve. [Last cited on 2013 Apr 15]. Available from: www.education.com/reference/article/itemanalysis
16. Scantron Guides - Item Analysis, adapted from Michigan State University website and Barbara gross devils tools for teaching. [Last cited on 2013 Apr 13]. Available from: www.freepdfdb.com/pdf/item-analysis-scantron
17. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educ Measure*. 2005:3–13.