Item Analysis of Multiple Choice Questions in Pharmacology

Imtiaz Uddin[1], Iftikhar Uddin[2], Izaz Ur Rehman[3] Muhammad Siyar[4] , Usman Mehboob[5].

ABSTRACT

Background: MCQs type assessment in medical education is replacing old theory style. There are concerns regarding the quality of the Multiple Choice Questions.

Objectives: To determine the quality of Multiple Choice Questions by item analysis.

Material and Methods: Study was a cross sectional descriptive .Fifty Multiple Choice Questions in the final internal evaluation exams in 2015 of Pharmacology at Bacha khan Medical College were analyzed. The quality of each Multiple Choice Questions item was assessed by the Difficulty index (Dif.I), Discriminative Index (D.I) and Distracter Efficiency (D.E).

Results: Multiple Choice Questions that were of moderate difficulty were 66%. Easy were 4% and high difficulty were 30%.Reasons for high difficult Multiple Choice Questions were analyzed as Item Writing Flaws 41%, Irreverent Difficulty 36% and C2 level 23%. Discrimination Index shows that majority of MCQs were of Excellent Level (DI greater than 0.25) i.e 52 , Good 32% . (DI=2.15-0.25), Poor 16%. MCQs Distracter Effectiveness (DE)= 4, 3,2,1 were 52%, 34%, 14%, and 0% respectively.

Conclusion: Item analysis gives us different parameters with reasons to recheck MCQ pool and teaching programme. High proportions of difficult and sizable amount of poor discriminative indices MCQs were the finding in this study and need to be resolved

Key Words: *Item analysis, MCQs Quality, Evaluation, pharmacology.*

INTRODUCTION

It is now well known that Multiple Choice Questions (MCQs) are gradually replacing the old essay style in medical examinations. Quality MCQ making is very difficult technically and time consuming[1].If the MCQs are properly constructed they can assess higher level of thinking skills. The cueing by the students is also minimize[2].If the quality of MCQs is ensured they also have good validity and reliability[2].Good quality MCQs mean that they should be having evidences for validity.Item analysis provides an evidence for validity, is a process based on the responses of the students. In this process quality of each MCQ item is assessed by three indices, Difficulty Index (Dif I), Discriminative Index (DI) Distracter Efficiency (DE) and' by Reliability.

Dif I show the level of difficulty of an MCQ. The higher is the (Dif .I) value the easier is the question. It is recommended that most of the MCQs in an examination should be of moderate

1. Department of Pharmacology, Bacha Khan Medical College, Mardan

2. Department of Community Medicine, Bacha Khan Medical College, Mardan

3. Department of Physiology, Saidu Medical College, Swat

4. Department of Bio chemistry, Bacha Khan Medical College , Mardan

5. Institute of Health Professional Education, Khyber Medical University Peshawar.

Correspondence: Dr. Imtiaz uddin, Assistant Professor
Department of Pharmacology Bacha Khan Medical College Mardan
Email: dr.imtiazuddin@gmail.com
Cell 03329485318.

level.i.e Dif is 0.4-0.8. Only 5% of the items are required to be in difficult range.[2.] According to the literature the Ideal Difficulty Index for an MCQ having five responses is 0.7 and difficult items showed be review for the reasons[3].However the guidelines about the qualitative interpretation of difficult items are also insufficient. There is a need that the difficult items should be reviewed and the reasons for the difficulty should be elaborated.

It is the power of right option to discriminate high performers from low performers. Various calculations have been used traditionally to find item discrimination by comparing responses of high and low scoring students to the total test score[4].The value of DI ranges between -1.0 to + 1.0.If discriminative index is negative for an item key may be wrong ,there may be presence of Item Writing Flaws or the course content is not properly taught or learnt[5]A value of greater than 0.15 Discriminative index has been suggested by a body of literature as the evidence of validity of an MCQs [2,6,7,]. If the value of DI for an MCQ is less than 0.15 up to negative range a review of item has been recommended to find out the reason [7]. Most of the studies have calculated the mean value of DI but the percentage of different reasons leading to Poor negative DI has not been clarified. They have not elaborated peer review needed for low DI items.

All the wrong options of an MCQs should be distracting the examinee to certain degree. The effectiveness of each wrong option of an MCQ to act as a distracter is called Distracter Index. Generally, Distracter Index for a wrong option is classified as non-functioning if it is selected by <5% and plausible if selected by >5% of the students[11] Distracter Efficiency is the total number of plausible options in an MCQ. Hence the maximum number for five option MCQ will be 4. There has been widespread body of research on the optimal number of distracter however there has been less focus on the status of plausibility of distracter[8,9]. Analysis of distracters plausibility is recommended to identify their errors so they may be replaced or removed.[11]

Reliability is the degree to which an assessment consistently produce similar results. The more reliable is an exam the greater is the confidence that the result will be the same if the exam is re-administer[10,12]. The reliability is measure by a correlation with 1.0 being ideal reliability (at least above 0.85).There are number of factors that affect reliability of the test[13] e.g. the length of a test, wider contents coverage ,environmental errors such as noise, temperature, heat, timings ,the homogeneity of examinee, the quality of test items

There has been little information related to the issue of reliability of MCQs in medical literature.
In order to find out the reasons of the above mentioned gaps this study has been conducted.

MATERIALS AND METHODS
It was a quantitative, cross-sectional descriptive study. The study was conducted in the Department of Pharmacology, Bacha Khan Medical Collage Mardan. The study was completed in six months i.e. Sep 2016 to Jan 2017.

Study population: All the MCQs in the subject of pharmacology in the internal assessments of BKMC. Mardan were the study population MCQs in pharmacology in the session 2015 internal exams fulfilling the criteria.

Inclusion Criteria:All the MCQs in the internal exams for the session 2015 were included in the study.

Exclusion criteria : Those response sheets were excluded from item analysis that were having more than 30% of the MCQs unmarked.

Ethical Review Board: The study was conducted after approval from Institutional Ethical Review Board.
Data Collection Procedure
a)   **The difficulty Index (Dif I)** It was classified as Easy for an item if corrected by >80%, Moderate corrected by 30% - 80% and Difficult if corrected by <30% of the examinees. The difficult items were reviewed by a panel of five examinees who attended the test. They were selected of having a varied level of academics. The panel concluded a reason of difficulty for each item. The Table-1 shows the details of the above procedure
b)   **Discrimination Index (DI)** It was calculated by the following formula
$$DI = 2*[(H-L)/N]\,^3$$
H = *number of correct responses in the higher one third group.*
L = *are the number of correct responses in the lower one third group.*
N = *total number of the students in both high and low groups.*
The DI was classified for an item as Poor if less than 0.15, Good if it is ranges from 0.15 to 0.24 and Excellent if it is greater than 0.25. The items that were having poor discriminative index was reviewed by two assessors to find the reasons for low DI. The details of the above information have been shown in Table-1

c)   **Distractor Efficiency:**Each MCQ item was rated according to the number of plausible distracters which was classified as non-functioning if selected by <5% and plausible if selected by =5%.An item having all the four distracters plausible was rated as 4 and so on.
d)   **Reliability:**
Reliability of the exam was found by KR20 using following formula [18.]

$$r = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{K} p_i q_i}{\sigma_X^2}\right]$$

The formula for KR-20 for a test with *K* test items numbered *i*=1 to *K* where p_i is the proportion of correct responses to test item *i*, q_i is the proportion of incorrect responses to test item *i* (so that $p_i + q_i =$ 1), and the variance for the denominator is

$$\sigma_X^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}.$$

where *n* is the total sample size.

*Panel review is done to find out reasons e.g. Item Writing Flaws, High level of Cognition, wrong key, ambiguity in language, insufficient teaching.

RESULTS
The finding shows that items of moderate difficulty were 66% Easy items were 4% and difficult MCQs (having low difficulty index) that needed preview were 30%**.**The review of the difficult items came to the following conclusion as 41% high difficult items were Item Writing Flaws related to Irreverent Difficulty, 23% were of C2 level and in 36% the difficult Items were due to insufficient teaching and learning.

The finding of Discrimination Index showed that majority of MCQs were of Excellent Level. 52 %.32% were good and Poor were 16% as shown in Figure. 1
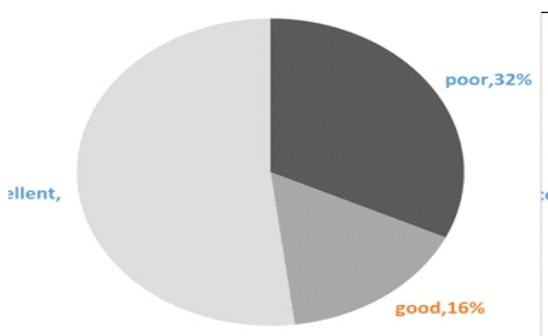

Fig 1 Percentages of Discriminative Indices

Panel review showed that reason for the Poor Discrimination Index was high level of Cognition (69%) and, irrelevant Difficulty type Item Writing Flaws (31%)

Relationship between Diff I and D I was found to be $R2=0.394$ not significant.
The evaluation of the items showed that most of the MCQs have Plausible Distracter. Distracter Effectiveness finding for DE= 4, 3, 2,1 were 52%, 34%, 14%, and 0% respectively.

Using the above mentioned formulae the Reliability was 0.62.

DISSCUSSION
Difficult items in this study were 30% while it is recommended that it should be up to 5%[6] .Previous studies have shown most of the items in moderate range[6,2] . Some of the previous studies have also shown a certain type of association between low Dif I and poor DI[14] In this study no such association was found between Dif I and DI.. So from this findings it can be conclude that poor DI is not a reason for low Dif I. This study through panel discussion found the reasons for low Dif I as Item Writing Flaws related to Irreverent Difficulty, C2 level and insufficient teaching and learning.

Discrimination Index (DI) was 52% excellent, 16% Good and 32% Poor. Previous studies have also shown similar type of result. [2,11,13]

Reasons for poor DI were C2 level of Cognition 68.7%, Irrelevant Difficulty 31.3% and wrong key 0% .

It can be concluded from the above findings that in present conventional teaching system if the items are generated at higher level of cognition they may lose the evidence for validity so the instruction methods need to be in alignment with assessments.

Table:1. Interpretation of the values of Difficulty Index and Discrimination Index[3]

Index	Cut off points	Interpretation	Action
Difficulty Index	<39	Difficult	PanelReview *
	40-80	Moderate	
	>80	Easy	
Discrimination Index	<0.15	Poor	PanelReview *

This study showed that most of the distracters were functional (Plausible) as DE= 4 were in 52% of the cases. This is in sharp contrast with the previous studies [2, 7,13, 15,] . Those studies showed large number of nonfunctional distractors. As a correlation was found between Dif I and DE so it can be concluded as a reason. All the items having four plausible distracters (DE=4) were having low Difficult Index. Most of the items having three plausible distracters were in difficult to moderate range.

Reliability was 0.62. This reliability is low in comparison to the previous studies [16, 17]. The reason may be attributed to the high proportion of items having low DI and less number of items with moderate difficulty.

The main limitation of this study was that it has been performed only in one institute and in a single subject. If same type of study is performed in multiple centers and in many subjects the scope of this study can be increased.

CONCLUSION

It can be concluded from this study that through panel discussion we can seek the reasons for Dif I. Keeping our present conventional system if items are generated at high level of cognition they may lose evidence of validity so instructional methods need to be in alignment with assessment. Plausibility of distracters increase with difficulty of items.

As the finding of this study shows high proportion of difficult and Poor DI items therefore the process of item analysis should be routinely performed in medical colleges for the improvement in the standard and quality of MCQs.

REFERENCES

1. Ali SH, Carr PA, Ruit KG. Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. Journal of the Scholarship of Teaching and Learning. 2016 Feb;16(1):1-4.
2. Elfaki OA, Bahamdan KA, Al-Humayed S. Evaluating the quality of multiple-choice questions used for final exams at the Department of Internal Medicine, College of Medicine, King Khalid University. Sudan Med Monit 2015;10:123-7
3. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. International Journal of Applied and Basic Medical Research. 2016 Jul;6(3):170-78.
4. Balaha MH. Simplified Guidelines for Multiple-Choice Question Writing to Increase Faculty Compliance and Ensure Valid Student Results. Education in Medicine Journal. 2019 Dec 1;11(4) 131-139.
5. Sim S M, Rasjah R I, Relationship between item difficulty and discrimination indices in true/false type multiple choice questions of a Para-clinical multidisciplinary paper. Ann Acad Med Singapore 2006;35:67-71
6. Hingorjo MR, Laleel F. Analysis of One -Best MCQs: the Difficulty Index, Discrimination Index and Distracter Efficiency. J Pak Med Assoc 2012;62:142-147
7. Chiavaroli N, Familiar M. When majority doesn't rule: The use of discrimination indices to improve the quality of MCQs. Beej 2011;17. Available from:http://www.bioscience.heacademy.ac.uk/journal /vol17/beej-17-8.pdf. [Last retrieved on 2017 Feb 4].
8. Mukhopadhyay M, Bhowmick K, Chakraborty S, Roy D, Sen KP, Chakraborty I. Evaluation of MCQs for judgement of higher levels of cognitive learning. Gom J Med Sci 2010;45-48.
9. Haladyna TM, Downing SM. How many options is enough for a multiple choice test item? Educ Psychol Meas 1993;53:999-1010
10. Haladyna TM, Downing SM. Validity of a taxonomy of multiple choice item-writing rules. Appl Meas Educ 1989;2:51-78
11. Back Hoff E, Larrazolo N, RosasM . The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination Elect J Educ Res 2012;2:9-14.
12. Linn RL, Gronlund NE. Measurement and Assessment in Teaching. New Jersey: Prentice-Hall Inc; 2000;1:574-80.
13. 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.
14. Carroll RG. Revaluation Vignette-type, examination items for testing medical physiology. Am J Physio 1993;264:S11-5.
15. Tarrant M, WareJ, Mohammed AM. An assessment of functioning and nonfunctioning distracter in multiple choice questions: Inter J Heal Sci & Res. 2009 ;9:40.
16. Fallatah HI, Tekian A, Soo Park Y and Shawa LA. The validity and reliability of the sixth-year internal medical examination administered at King Abdulaziz University Medical College. Medical education 2015;15:10
17. Auewarakul C[1], Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. Med Educ.2005;39(3):276-83
18. Downing SM. Reliability: on the reproducibility of assessment data. Medical education. 2004 Sep;38(9):1006-12.