



# Item analysis: An evaluation of multiple choice questions in Physiology examination

Anjali Nilkanthappa Shete, Afshan Kausar, Kunda Lakhkar, S. T. Khan

## ABSTRACT

**Background:** Item analysis is the process of collecting, summarizing and using information from student's responses to assess the quality of test items. Difficulty index and discrimination index are two parameters which help to evaluate the standard of multiple choice question (MCQ) used in an examination. **Materials and Methods:** An internal examination in Physiology was conducted at the end of one term. 148 students, out of 150, attended the exam. A total of 40 MCQ were given and analyzed. The data was analyzed by Microsoft Excel and recorded as mean  $\pm$  standard deviation. The relationship between the difficulty index (P) and discrimination index (D) for each test item was determined by Pearson correlation analysis. **Results:** The mean difficulty index score was  $0.54 \pm 0.26$  and mean discrimination score was  $0.21 \pm 0.14$ . On an average, about 10 of the MCQs were easy ( $P \geq 70\%$ ) while 12 were difficult ( $P \leq 30\%$ ) and the remaining 18 items were in acceptable range ( $P = 30-70\%$ ). In all, 2 items showed negative discrimination and 13 items exhibited poor discrimination. The remaining 25 items were in the range of acceptable to excellent discrimination. The discrimination index exhibited slight positive correlation with difficulty index ( $r = 0.3076$   $P = 0.05$ ). However, it was not statistically significant. The maximal discrimination ( $d = 0.55$ ) was observed with easy/difficult items ( $P = 0.6$ ). **Conclusion:** The majority of the items were in acceptable range as far as difficulty and discrimination indices were concerned. Moderately easy/difficult items had the maximal discriminative ability. Too easy and too difficult items gave poor discrimination index.

**KEY WORDS:** Item analysis, Mcqs, difficulty index, discrimination index

Department of Physiology,  
Government Medical  
College, Aurangabad,  
Maharashtra, India

### Address for correspondence:

Anjali Nilkanthappa Shete,  
Department of Physiology,  
Government Medical  
College, Aurangabad,  
Maharashtra, India.  
E-mail: dranju01@yahoo.  
com

**Received:** May 30, 2015

**Accepted:** August 20, 2015

**Published:** October 26, 2015

## INTRODUCTION

The educational objectives in medicine as well as in other discipline are generally allotted to three domains-cognitive, psychomotor, and affective. Hence, medical examination should be designed to answer whether an undergraduate has achieved the above educational objectives by answering the following three questions:

1. What does he know? (cognitive domain)
2. What can he do? (psychomotor domain)
3. What sort of person he is? (affective domain) The current medical system still could not answer these questions completely [1].

Designing well multiple choice question (MCQ) is a complex and time-consuming process. After construction and assessment of MCQs; they need to be tested for the standard or quality. The single best response type of MCQ are designed to assess knowledge [2]. They have the advantage of sampling broad domains of knowledge effectively and reliably. This characteristic of MCQ gives some reliability for assessment. If carefully constructed, MCQs test higher order thinking

skills [3,4]. Therefore, MCQs remain a useful assessment instrument; despite some limitations and objections.

Item analysis is a process which examines student's responses to individual test items to assess the quality of these items and quality of test as a whole. It is of great help in improving the quality of items which may be used again in subsequent tests. It also improves the skill in the construction of test items and also helps identify course content which needs greater emphasis or clarity. It provides feedback to teachers to install changes in the standard of teaching. The item statistics can help find out poor items which need improvement or deletion. It allows any aberrant items to be given attention and reconstructed [5-7].

The Medical council of India as required by the regulation on the Graduate Medical Education 1997, made it mandatory for all medical colleges to establish Medical Education Units to enable faculty members to avail modern medical education technology for teaching. To boost this activity, MCI has been conducting faculty development programs through selected regional centers. These centers have trained manpower in

teaching modern medical education technology (MET) [8]. Item analysis is a part of MET training.

With this background, the objectives of our study were to analyze the quality of MCQs of Physiology examination by doing the Item analysis and to determine the relationship between the difficulty and discrimination indices of these MCQs.

### MATERIALS AND METHODS

At the end of one term, the first MBBS students were given an examination on Physiology. 148 students appeared out of 150 students. They were given 40 MCQs having half mark each for correct answer. No negative marking was there for the wrong answer. All MCQs were single best response type with four distracters. The MCQs were constructed by all teachers in the department.

The result of student’s performance in these MCQ test was used to determine the level of difficulty and power of discrimination using Microsoft office Excel. The steps for item analysis were scoring of the whole test for all students rank students in order of merit based on test scores:

1. Top third were taken as high achievers (h) and bottom third (l) as low achievers

2. Table 1 was prepared for each item to get the value of h, and the calculations were made using the following formulae from the books of Medical Education [9,10].

$$\text{Difficulty index (P)} = \frac{h + l}{n} \times 100$$

$$\text{Discrimination index (D)} = \frac{h - l}{n} \times 2$$

Where;

h = number of students answering correctly in high achievers group.

l = number of students answering correctly in the low achievers group.

n = total number of students in both groups including non-responders.

3. Interpretation

Difficulty index (P):  $P \leq 30\% \rightarrow$  difficult

$P = 30-70\% \rightarrow$  Acceptable

$P \geq 70\% \rightarrow$  Easy

Discrimination index (D):  $D = \text{Negative} \rightarrow$  Defective item/wrong key

$D = 0-0.19 \rightarrow$  Poor discrimination

$D = 0.2 - 0.29 \rightarrow$  Acceptable discrimination

$D = 0.3 - 0.39 \rightarrow$  Good discrimination

$D \geq 0.4 \rightarrow$  Excellent discrimination

Hence, the higher the difficulty index value; the lower is the difficulty and the lower the difficulty index value; the greater is the difficulty of an item. For discrimination; higher the index, better the item can discriminate among those students with high test scores and those with low ones [9-11].

**Table 1: Difficulty index and discrimination index of MCQS**

Item analysis parameters	Mean±SD	Range
Difficulty index	0.54±0.26	0.07-0.97
Discrimination index	0.21±0.14	-0.05-0.55

MCQS: Multiple choice questions, SD: Standard deviation

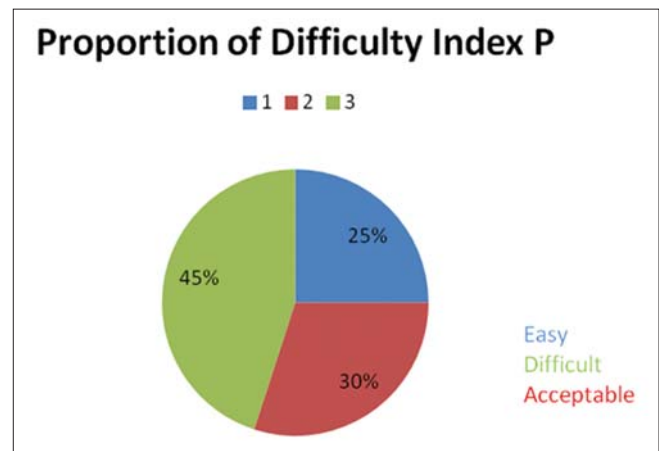
### Statistical Analysis

The data were reported as % and mean ± standard deviation of all items. The relationship between the item difficulty index and discrimination index values for all items was determined using Pearson correlation analysis. P value of < 0.05 was considered to indicate statistical significance.

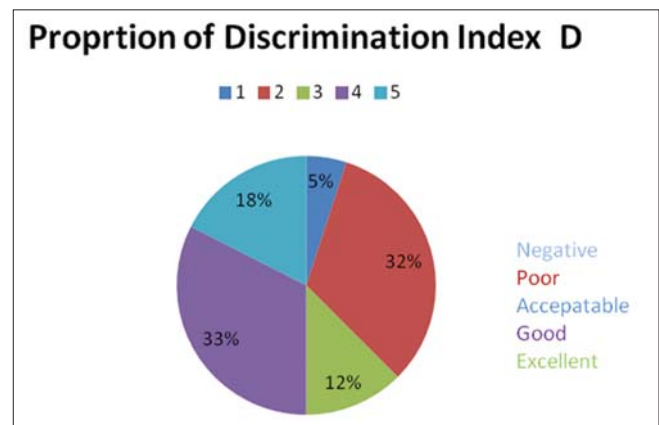
### OBSERVATIONS AND RESULTS

As seen in Table 1, mean difficulty index (P) was  $0.54 \pm 0.26$  while mean discrimination index (D) was  $0.21 \pm 0.14$ . Figure 1 shows that out of total 40 items, difficulty indices of 10 MCQ items were easy ( $P \geq 70\%$ ) while about 12 MCQ were difficult ( $P \leq 30\%$ ), and the remaining 18 of the items were within acceptable range ( $P = 30-70\%$ ).

As seen in Figure 2, the discrimination index (D) for 40 items; 2 items showed negative discrimination, 13 were having poor discrimination ( $D < 0-0.19$ ) while only 5 were having excellent discrimination ( $D > 0.4$ ). The remaining 20 were acceptable and good; out of which 13 items were having acceptable ( $D = 0.2-0.29$ ), and 7 items were having good discrimination ( $D = 0.3-0.39$ ).



**Figure 1:** Proportion of difficulty index



**Figure 2:** Proportion of discrimination index

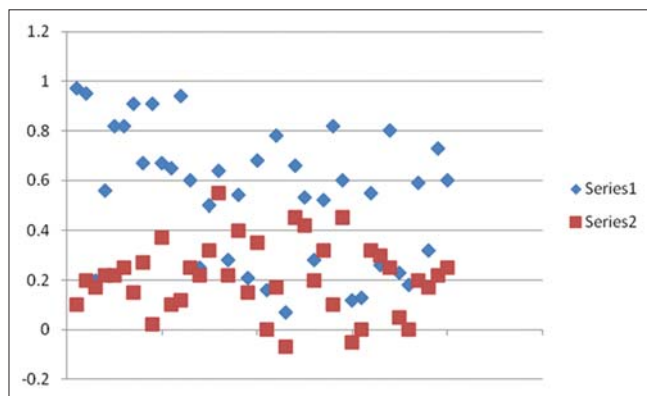
The scattered diagram Figure 3 represents the relationship between the difficulty index and the discrimination index of 40 MCQs. The discrimination index correlated positively with the difficulty index ( $r = 0.3076, P = 0.053$ ) which is not significant statistically. The maximal discrimination ( $D = 0.55$ ) had moderate easy/difficult items ( $P = 0.64$ ). It is seen that in two of the items had negative discrimination indices value ranging from  $-0.05$  to  $-0.07$  with the corresponding difficulty index between  $0.07$  and  $0.12$ . This may be due to faulty items or a wrong key.

## DISCUSSION

The effective measurement of knowledge acquired is an important component of medical education. MCQs are useful assessment tools in measuring factual recall, and carefully constructed MCQs can test higher order of thinking skills which is very important for a medical graduate [3,4]. The method of assessment should be regularly evaluated. Fowell *et al.* have stressed the importance of this step of assessment which is often omitted [12]. It is important to evaluate MCQs items to see how effective they are in assessing the knowledge of students. Items that discriminate poorly should be reviewed for possible corrections and reconstruction or deletion. Some basic forms of item analysis may be carried out routinely and the data generated should be used regularly to test the quality of the questions or for the development of MCQs for the subsequent tests.

In this study, the wide scatter of difficulty and discrimination indices was observed indicating some guessing, may be due to no negative marks allotted to wrong answers. Same observations were reported by Sim and Rasiah [6] in their study on true/false questions and MCQs in the para-clinical examination. Mitra *et al.* in 2009 [11] also had similar results in their study.

In present study, 18 of the items had acceptable difficulty indices ( $P = 30-70\%$ ), 10 were easy ( $P > 70\%$ ) while 12 of the items with  $P < 30\%$  were difficult. This could have been due to poor understanding of difficult topics, ambiguity in wordings of the questions or inappropriate key or personal variation in forming



**Figure 3:** Corelation of difficulty and discrimination index. Series 1 : Difficulty index, Series 2: Discrimination index

the MCQs and may also be a clue to variation in students intelligence level. The discrimination index ( $D$ ) serves as an effective feedback to teachers about the quality of each item. Items with poor discrimination should be reviewed. According to Brown [13] and Crocker and Algina [14];  $D > 0.2$  is acceptable and able to discriminate between good and weak students. The present study showed that 5 of the items had  $D > 0.4$  which is excellent discrimination, 20 of items showed good and acceptable discrimination  $D > 0.2-0.29$ , 13 items were having poor discrimination  $D < 0-0.19$ . In all, only 2 of the items had negative discrimination.

A similar type of study reported by Ho *et al.* [1] showed that too easy or too difficult items discriminate poorly. However, the correlation between the two indices was not done. A study done Pande *et al.* [15] reported positive correlation in difficulty and discrimination indices. Studies by Sim and Rasiah [6] and Mitra *et al.* [11] showed that the discrimination index correlated poorly with the difficulty index. The correlation signified that with the increasing difficulty index values, there was a decrease in discrimination index indicating that low-performance students were more likely to get the correct answers. As the items got easier, the level of discrimination index decreased consistently [11].

Very difficult and very easy items need to be reconstructed and reevaluated. An item analysis will serve as a helpful tool to generate MCQ banks at departmental and university levels. Administration of an objective test and use of item analysis at the end of the period, sometimes even as small as a single lecture, has great advantages for the teacher. It enables teachers to get active feedback from the students and determine areas which require emphasis, reinforcements, or an alteration in teaching methodology. Although every aspect of an instructional exercise cannot be reduced to MCQ; use of items frequently during classroom teaching especially in problematic areas helps the teacher in improving his and his student's performance. Usually, items which have a good positive discrimination and moderate difficulty are chosen. Teachers must aim at getting high facility values and low discrimination indices as the aim of teaching is not to distinguish between good and bad students but to ensure that all students have learnt the lesson correctly [9].

## CONCLUSION

In the present study, the majority items fulfilled the criteria of acceptable difficulty and good discrimination, which means the MCQs selected were of good quality. Moderately easy/difficult had maximum discrimination ability. Very easy and very difficult items displayed poor discrimination. Items with negative and poor discrimination will be reviewed, reconstructed, and added to the departmental MCQ bank.

## ACKNOWLEDGMENT

I would like to thank all the teachers in Physiology Department for contributing MCQs and their valuable support.

## REFERENCES

1. Ho TF, Yip WC, Tay JS. The use of multiple choice questions in medical examination: An evaluation of scoring and analysis of results. *Singapore Med J* 1981;22:361-7.
2. Skakun EN, Nanson EM, Kling S, Taylor WC. A preliminary investigation of three types of multiple choice questions. *Med Educ* 1979;13:91-6.
3. Norman G. The program for educational development. *Evaluation Methods A Resource Handbook, Multiple Choice Questions*. Ch. 4.1. Hamilton, Canada: McMaster University; 1995. p. 47-54.
4. Peitzman SJ, Nieman LZ, Gracely EJ. Comparison of "fact-recall" with "higher-order" questions in multiple-choice examinations as predictors of clinical performance of medical students. *Acad Med* 1990;65:S59-60.
5. Zubairi AM, Kassim NL. Classical and Rasch analysis of dichotomously scored reading comprehension test items. *Malays J ELT Res* 2006;2:1-20.
6. Sim SM, Rasiah RI. 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.
7. Boopathiraj C, Chellamani K. Analysis of test items on difficulty level and discrimination index in the test for research in education. *Int J Soc Sci Interdiscip Res* 2013;2:189-93.
8. Faculty Development Program. Available from: <http://www.mciindia.org/information desk/for colleges/faculty development program>.
9. Ananthkrishnan N. The item analysis. *Medical Education Principles and Practice*. 2<sup>nd</sup> ed. Pondicherry: JIPMER; 2000. p. 131-7.
10. Tejinder S, Piyush G, Daljit S. *Principles of Medical Education*. 3<sup>rd</sup> ed. India: Jaypee Brothers, Medical Publishers Pvt. Limited; 2009. p. 70-7.
11. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in Type A multiple choice questions of preclinical semester 1 multidisciplinary summative tests. *Jpn Soc Mech Eng* 2009;3:2-7.
12. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: The missing link? *Med Educ* 1999;33:276-81.
13. Brown FG. *Principles of Educational and Psychological Testing*. 3<sup>rd</sup> ed. New York: Holt, Rinehart and Winston; 1983.
14. Crocker L, Alagina J. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston; 1986.
15. 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;7:45-50.

© SAGEYA. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.

**Source of Support: Nil, Conflict of Interest: None declared.**