ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Item-analysis of Multiple Choice Questions: A Pilot Attempt to Analyze Formative Assessment in Pharmacology.

Preethi J Shenoy*, Vinaykumar Sayeli, and Rashmi R Rao.

Department of Pharmacology, Kasturba Medical College, Mangalore, Manipal University, Manipal, Karnataka, India.

ABSTRACT

The aim of the study was to assess the quality of multiple choice questions, for creating a viable question bank for future use. The study was conducted in the department of Pharmacology. The marks obtained by the second year MBBS students in the MCQ tests during Pharmacology formative examinations were analyzed. Each correct response was awarded 1 mark and each incorrect response was awarded 0, with no negative marking. Post validation of the paper was done by item analysis. Each item was analyzed for difficulty index, discrimination index and distractor effectiveness. Difficulty index of 24 (40%) items was in the acceptable range (p value 30-70%), 21 (35%) items were too easy (p value >70%) and 15 (25%) items were too difficult (p value <30%). Discrimination index of 12 (20%) items was excellent (d value>0.35), 36 (60%) items was good (d value 0.20-0.34) and 12 (20%) items were poor (d value<0.2%). A total of sixty items had 180 distractors. Amongst these, 60 (33.3%) were nonfunctional distractors, 63 (35%) were functional distractors and 57 (31.7%) had nil response i.e. not attempted by any student. On the basis of non-functional distractors, distractor effectiveness of each item was assessed. This study inferred that there is a genuine need for regular item analysis of multiple choice questions to generate a question bank with an intention to achieve uniform, reliable and valid assessment. Questions with average difficulty and high discriminating power with functional distractors should be incorporated into future tests to improve the validity and reliability of MCQ as a mode of formative assessment.

Keywords: multiple choice question, difficulty index, discrimination index, distractor efficiency, item analysis

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Assessment is an important component of a teaching-learning curriculum. Good assessment reflected by its validity, reliability and acceptability to students and faculty can prove to improve the educational impact of student evaluation. A significant application of formative assessment is to continuously monitor learning activities and give appropriate feedback to students and teachers [1]. Multiple choice questions (MCQ) is a frequently used mode of student assessment in medical education especially due to its inherent advantage of being able to test large number of students within a short time interval and minimal human intervention [2]. However MCQs have been heavily criticized of having taken the context out of real world solutions as various academicians feel that a different cognitive process is involved in proposing a solution versus selecting a solution from a set of alternatives. MCQs which are based on factual recall may fail to test the students' higher levels of cognition, problem solving ability and organizational competency. However, this is a flaw on the side of the exam setter rather than the inherent weakness of MCQ. MCQs designed to test objectively and measure knowledge, comprehension, application, analysis and evaluation can nullify most of these criticisms [2].

According to Tarrent and Ware flawed MCQ items affect the performance of high-achieving students more than borderline students [3]. Designing well-structured MCQs can address the concern of the students of getting an acceptable average score and that of the faculty to have an appropriate spread of the marks [4]. Item analysis has proven to be an important step in the development of MCQ as an assessment strategy [5]. It is the phase that helps us identify an item that is either too easy or too difficult for the student. This process also helps in detecting items that fail to discriminate between high and low scoring students. It helps in contributing to the fairness of the test along with identifying content areas that maybe problematic for students. It is the process of collecting, summarizing and using information from students' responses to assess the quality of test items [6].The item statistics can help to determine those MCQs which are good and those that need improvement or deletion from a question bank. It helps us to identify aberrant items and review them [7].

Assuming that the overall quality of the test is dependent on the quality of items used, item statistics are used to assess the performance of individual test items. Statistical methods are used for these purposes. The common methods adopted in an item analysis are the facility value, discrimination index and distractor efficiency [8]. Facility value, Difficulty index (p-value), also called ease index, describes the percentage of students who correctly answered the item. It ranges from 0 - 100%. The higher the percentage, the easier the item. The recommended range of difficulty is from 30 - 70%. Items having p-values below 30% and above 70% are considered difficult and easy items respectively [9]. Discrimination index (Item discrimination) (DI), describes the ability of an item to distinguish between high and low scorers [10]. It ranges between -1.00 and +1.00. It is expected that the high-performing students select the correct answer for each item more often than the low-performing students. If this is true, the assessment is said to have a positive DI (between 0.00 and +1.00), indicating that students who received a high total score, chose the correct answer for a specific item more often than the students who had a low overall score. If, however, the low-performing students got a specific item correct more often than the high scorers, then that item has a negative DI (between -1.00 and 0.00) [11]. Items with negative DI are not only useless, but actually serve to decrease the validity of the test.

Analysis of distractor is another important part of item analysis. The distractors are important components of an item, as they show a relationship between the total test score and the distractor chosen by the student. Student's performance depends upon how the distractors are designed [12]. Distractor efficiency is one such tool that tells whether the item was well constructed or failed to perform its purpose. Any distractor that has been selected by less than 5% of the students is considered to be a non-functioning distractor (NF-D) [13]. Ideally, low-scoring students, who have not mastered the subject, should choose the distractors more often, whereas, high scorers should discard them more frequently while choosing the correct option. By analyzing the distractors, it becomes easier to identify their errors, so that they may be revised, replaced, or removed [14].

With this background the present study is a pilot attempt at validating the MCQs used in a formative assessment of Pharmacology in second year MBBS students with the aim of creating a viable question bank for future use.

ISSN: 0975-8585

MATERIALS AND METHODS

The marks obtained by the second year MBBS students in the MCQ tests during three (First, second and third sessional) Pharmacology formative examinations were analyzed. It comprised of 60 "single response type" MCQs. The examination time was 60 minutes and marks allotted were 60. All MCQs had single stem with four options/responses including, one being correct answer and other three incorrect alternatives (distractor). Each correct response was awarded 1 mark and each incorrect response was awarded 0, with no negative marking. A list was generated with the marks of the student scoring highest marks at the top followed by other scores arranged in a descending fashion. The whole list was then divided into two equal groups: the group of high achievers (Higher group or HG) and low achievers (Lower groups or LG).

Item statistics: The various indices were analyzed using the following formulae:

Facility value (FV):

$$FV = \frac{HG + LG}{N} \times 100$$ where N= total number

Discrimination index (DI):

$$DI = \frac{2 \times (HG - LG)}{\text{Total number}}$$

Items with p-value between 30 - 70 and DI > 0.24 were considered as 'ideal'. NF-Ds will be those selected by less than 5% of students.[15] Distractor efficiency were ranged from 0 - 100% and were determined on the basis of the number of NF-Ds in an item.

Four NF-D: DE = 0%; 3 NF-D: DE = 25%; 2 NF-D: DE = 50%; 1 NF-D: DE = 75%; No NF-D: DE = 100%.

At the end of the item analysis report, test items were listed according to their degrees of difficulty (easy, medium, hard) and discrimination (good, fair, poor). These distributions were used to obtain a quick overview of the test and identify items not performing well and which can perhaps be improved or discarded. The acceptable MCQs thus collected were then be retained in a MCQ bank created in the Department.

RESULTS

Difficulty index or facility value

As shown in figure 1, the number of items having ideal facility value (p value 50-60%) was 24 (40%).

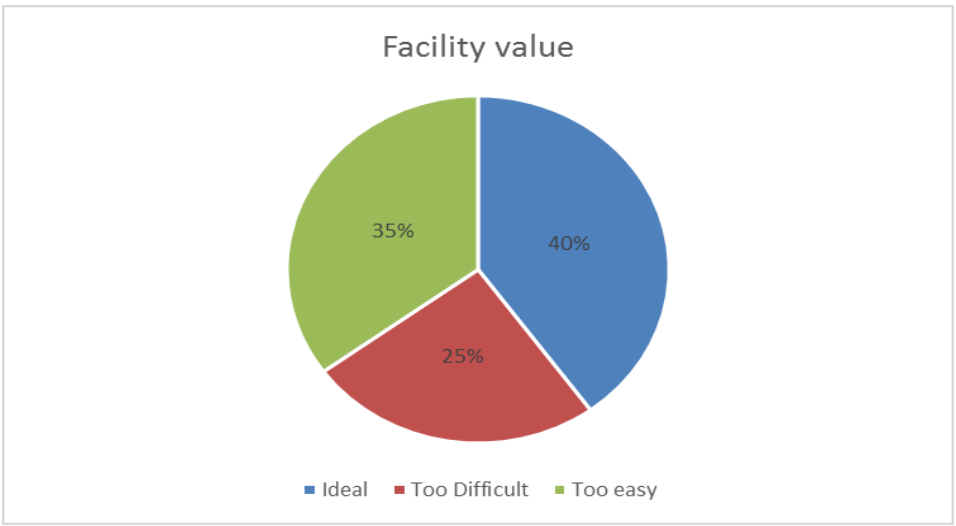

Figure 1: Difficulty index of MCQ items

ISSN: 0975-8585

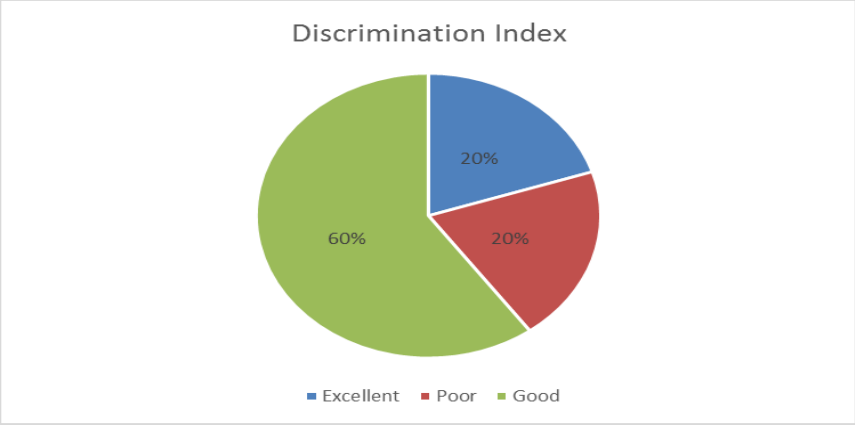

Figure 2: Discrimination index of MCQ items

Total number of items: 60			
Items with 0 NFD DE=100%	Items with 1 NFD DE =66.6%	Items with 2 NFD DE= 33.33%	Items with 3 NFD DE= 0%
21 (35%)	21 (35%)	15 (25%)	3 (5%)

Table 1: NFD and distractor effectiveness (DE) of each MCQ item

Discrimination index

Discrimination index of 12 (20%) items was excellent (d value>0.35), 36 (60%) items was good (d value 0.20-0.34) and 12 (20%) items were poor (d value<0.2%) as evidenced in figure 2.

Distractor effectiveness

A total of sixty items had 180 distractors. Amongst these, 60 (33.3%) were nonfunctional distractors, 63 (35%) were functional distractors and 57 (31.7%) had nil response i.e. not attempted by any student as depicted in table 1.

DISCUSSION

Effective measurement of knowledge is an important component of medical education. MCQ evaluation helps us to see how effective they are in assessing the knowledge of our medical students, and in predicting their total test scores. Each item (MCQ) while being used in assessment must be evaluated based on FV, DI, and DE because if an item is flawed then this itself becomes distracting and the assessment can be false. Too difficult items (DIF I ≤ 30%) will lead to deflated scores, while the easy items (DIF I > 60%) will result into to inflated scores and a decline in motivation. Too easy items should be placed either at the start of the test as 'warm-up' questions or removed altogether, similarly difficult items should be reviewed for possible confusing language, areas of controversies, or even an incorrect key.

In the present study, analysis of the **facility value** indicated that 40% of the questions were ideal i.e. their facility value was between 60-70%. 35% of the questions were too easy with a facility value of 70% which was higher than recommended limit of 10-20%. 25% of the questions were too difficult with a facility value of < 30% whereas the recommended limit is 20%.

Inclusion of very difficult items in the test depends upon the target of the teacher, who may want to include them in order to identify top scorers [11]. Facility value is relevant for determining whether students have learned the concept being tested. It helps to understand about the learning that happened in the classroom [16]. It also plays an important role in the ability of an item to discriminate between students who know the tested material and those who do not. The item will have low FV, if it is so difficult that almost everyone gets it wrong or guesses, or so easy that almost everyone gets it right. Knowledge of FV helps in better design of the paper [2].

On **discriminatory index (d),** there are only 4 questions (20%) to be considered as 'Excellent'; 12 questions (60%) are 'Acceptable' whereas, the rest 20% either need revision or to be discarded. DI of an item indicates its ability to differentiate between students of higher and lower abilities. It is obvious that a question which is either too difficult (done wrongly by everyone) or too easy (attempted correctly by everyone) will have nil to poor DI. Value of DI normally ranges between 0 and 1. The reasons for negative DI are often due to an ambiguous question or an answer key that was wrongly marked [2]. The difficulty and discrimination indices are often reciprocally related. However, this may not always be true. Questions having high p-value (easier questions), discriminate poorly; conversely, questions with a low p-value (harder questions) are considered to be good discriminators [17]. DI is thus used to identify flaws in the items, to provide improvement options, improve learning and identify misconceptions in learning [2].

Analyzing the distractors (incorrect alternatives) is done to determine their relative usefulness in each item. Items need to be modified if students consistently fail to select certain distractors. Such alternatives are probably implausible and therefore of little use as decoys. Therefore, designing of plausible distractors and reducing the NFDs is important aspect for framing quality MCQs. More NFD in an item increases DIF I (makes item easy) and reduces DE, conversely item with more functioning distractors decreases DIF I (makes item difficult) and increases DE. Higher the DE more difficult the question and vice versa, which ultimately relies on presence/absence of NFDs in an item. In the present study, 7 (35%) of the questions had items with either 0 or 1 NFD whereas 5 (25%) questions had 2 NFD and 1 question has 3 NFD with very poor distractive effectiveness. Flawed MCQs interfere with accurate, meaningful interpretation of test scores and negatively impact pass rate. Therefore, to develop reliable and valid tests, items are required to be flawless.

CONCLUSION

Item analysis of MCQ papers reveals existence of wide variation in difficulty level and their discriminative ability. There is a genuine need to review quality of contents. It is only through the repetitive process of item analyses and improvement that pedagogically and psychometrically sound tests can be developed [13]. MCQs may be validated each time before next examination for suitable rectification of weakness, besides enriching the question banks. This is bound to improve evaluation quality, bringing in radical changes in examination system in terms of uniformity and objectivity.

REFERENCES

[1] Gyata M, Varsha M, International journal of health science and research 2014;4 (7): 197-202.
[2] Ciraj AM. Multiple choice questions In: Principles of assessment in medical education. Jaypee Publications India, 2012, pp. 88-107.
[3] Tarrant M, Ware J, Med Educ 2008; 42: 198-206.
[4] Carroll RG, Am J Physiol 1993; 264: S11-5.
[5] Ebel RL & Frisbie DA, Essentials of educational measurement, Englewood, NJ, Prentice Hall.
[6] Mitra NK, Nagaraja HS, Ponndurai G et al, IeJSME 2009; 3(1): 2-7.
[7] Bharti N Karelia, Ajita Pillai, Bhavisha N Vegada. IeJSME 2013 7(2): 41-46.
[8] Thorndike, RL, The analysis and selection of test items: In: Problems in human assessment. New York: Mc Graw-Hill, 1967.
[9] Miller MD, Linn RL, Gronlund NE, Measurement and assessment in teaching. 10th ed. Upper Saddle River, NJ: Prentice Hall; 2009.
[10] Fowell SL, Southgate LJ, Bligh JG, Med Educ 1999; 33: 276-81.
[11] Hingorjo MR, Jaleel F, Journal of Pakistan Medical Association 2012;62(2):142-7.
[12] Dufresne RJ, Leonard WJ, Gerace WJ, Phys Teach 2002; 40: 174-80.
[13] Tarrant M, Ware J, Mohammed AM, BMC Med Educ 2009; 9: 40.
[14] Gronlund NE, Linn RL, Measurement and evaluation in teaching, New York, Macmillan publishing Co, 1990.
[15] Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T, Indian Pediatrics 1998; 35: 1103-5.
[16] Haladyna TM. Developing and validating multiple choice test items. 1999, Mahwah, NJ, Lawrence Erlbaum Associates.
[17] Carroll RG, Am J Physiol 1993; 264: S11-5.