Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2018, 10(9): 50-54

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Item Analysis to Identify Quality Multiple Choice Questions in Pharmacology Examination

Chaudhari Vijaya Laxman* and Moghe Vijay Vasant

Department of Pharmacology, SMBT IMS & RC, Dhamangaon, Maharashtra, India

---

ABSTRACT

Evaluation is an important component of a teaching learning curriculum. Multiple choice questions (MCQs) are used as an objective and reliable tool to evaluate learning performance of students. The aims of this study were to evaluate MCQs for developing a pool of valid items and to update question bank for designing question paper in future. The marks obtained by 122 second year MBBS students in total 88 MCQs during three internal theory examinations in Pharmacology were analysed. Each correct response was awarded 1 mark while incorrect response was awarded 0. A list was generated with the marks of the student scoring in a descending fashion. The list was then divided into two equal group as high achievers and low achievers. MCQ analysis was conducted based on difficulty index (DIF I), discrimination index (DI), and distractor efficiency (DE). DE of items with different values of DIF I and DI was further analyzed by SPSS with 95% level of significance using unpaired t- test. Total 88 MCQs and 264 distractors were analyzed. Means and standard deviations (SD) for DIF I, DI and DE were 58.15 ± 2.19%, 0.69 ± 0.03, and 19.68 ± 3.26%, respectively. The difference in DE among changing DI was statistically significant with t=9.61, df=82, p<0.0001. Item analysis is a simple and feasible method of assessing valid MCQs in order to achieve the ultimate goal of medical education.

Keywords: Difficulty index; Discrimination index; Distractor efficiency; Non-functional distractors

---

INTRODUCTION

Improving the health and the health care of the population is the principal objective of the medical education [1]. To fulfill this objective the assessment of medical students is a prerequisite. The multiple choice questions (MCQs) have become the essential part of such assessments in medical colleges. However, these MCQs must be pre-validated [2]. Therefore, Haladyna et al. had prepared the guidelines for framing the MCQs [3]. According to Gajjar et al. [4] a good MCQ truly checks the knowledge and is able to differentiate the students of different abilities, while Sharif et al. concluded that MCQ is an effective and reliable tool for determining the accomplishment of the Medical students [5].

Scheming good quality MCQs is a complex, challenging and time consuming process. Studies have reported that it is five times faster to revise MCQs that didn't useful, using item analysis, than working to replace it with a completely new question whose validity is unknown. New MCQ may have new problems [6]. After designing the MCQs need to be tested for the standard or quality. Item analysis studies the student answers to individual test items (MCQs) to evaluate the quality of those items and test as a whole [7]. It is very important and simpler procedure conducted after the examination that offers information regarding the reliability and validity of the MCQ test [8].

Although the fact that preparation of a good item is very much essential to provide a valid MCQ, it is hardly attempted by the examiners. Thus, the present study has been carried out with an objective to evaluate of MCQs to develop a pool of valid items and to update question bank for designing question paper as per the need of assessment in future.

METHODOLOGY

The present study was conducted in the Department of Pharmacology, SMBT IMS and RC, Dhamangaon, Igatpuri and commenced after the Institutional Ethics Committee approval (SMBT/IEC/2017/Project-59).The marks obtained by 122 second year MBBS students (batch Aug 2015-Dec 2016) in the Multiple Choice Question (MCQ) section during three internal theory examinations (First, second and third sessional) in Pharmacology were analyzed. The only inclusion criteria was students attended all the theory examination in Pharmacology at the same time. The examinations comprised of total 88 (28+28+32) "single response type" MCQs. All MCQs had single stem with four options comprising of one 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 the low achievers (Lower groups or LG). Each item was analyzed for:

Difficulty Index

The percentage of total number of students from both the groups (High Achievers and Low Achievers) opting for key (i.e. answering correctly) represents the difficulty index (denoted as 'p'). Difficulty Index (DIF I) or p value was calculated using the formula-

DIF I or p=H+L/N × 100

H=number of students answering the item correctly in the high achievers group.

L=number of students answering the item correctly in the low achievers group

N=Total number of students in the two groups (including non-responders)

In general, Items with difficulty index less than 30% were considered as difficult. If an item has a 'p' value between 30-60% it was considered as acceptable. Items with difficulty index greater than 60% were considered as easy.4

Discrimination Index

Denoted as 'DI' measures the ability of an item to discriminate between students. To calculate this index the numbers of responders to the 'key' was taken into account. Here the difference between the two groups was found out. The larger the difference between high achievers and low achievers, the greater will be the discrimination power of an item. Discrimination index (DI) value was calculated using the formula

DI=(H-L) × 2/N

Where the symbols H, L and N represent the same values as mentioned above.

The Discrimination index ranges from 0 to 1. An index value of +1 means the item has maximum discriminative power. An item having a discrimination index greater than 0.25 was considered as to have excellent discriminative power. An item having a discrimination index between 0.15 and 0.25 has acceptable discriminative power. An item having discrimination index 0 cannot discriminate between two (H and L) groups [4].

Distracter Effectiveness (DE) or Functionality

An item contains a stem and four options including one correct (key) and three incorrect (distracter) alternatives. Non Functional Distracter (NFD) in an item is the option, other than the key selected by less than 5% of students and functional or effective distracter is the option selected by 5% or more students. On the basis of number of NFDs in an item, DE ranges from 0 to 100%. If an item contains three or two or one or nil NFDs then DE would be 0, 33.3%, 66.6% and 100% respectively [4].

After completing the item analysis, MCQs were enlisted according to their degrees of difficulty (easy, excellent, good, and difficult) and discrimination (excellent, good, and 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 being retained in a MCQ bank created in the Department. Distractors efficiency (DE) of items with different values of DIF I and DI will be further analyzed by SPSS version 24.0 (IBM Corporation, Armonk, New York) with 95% level of significance using unpaired t- test.

RESULT

Total 88 MCQs and 264 distractors were analyzed. Means and standard deviations (SD) for DIF I (%), DI and DE (%) were 58.15 ± 2.19%, 0.69 ± 0.03, and 19.68 ± 3.26%, respectively (Table 1). Out of 88 items, 46 had "good to excellent" level of difficulty (DIF I=31-60%) and 82 had "good to excellent" discrimination power (DI≥ 0.15) (Tables 2 and 3). When these two were considered together, there were 46 items as ideal which could be included in question bank. Out of 264 distractors, 52 (19.7%) were NFDs present in 31 items (15 had 1, 11 had 2 and five had 3) with DE varying between 0 and 66.6%. Remaining 57 items had no NFDs with their DE being 100%. The 31 items with NFDs had mean values of DIF I and DI as 75.30% and 0.43, respectively (Table 4). The 57 items without NFDs had mean values for DIF I and DI as 48.82% and 0.83, respectively.

Table 1. Assessment of 88 items based on various indices amongst 122 students

Parameter	Mean	Standard Deviation (SD)
Difficulty index (DIF I) (%)	58.15	2.19
Discrimination index (DI)	0.69	0.03
Distractor efficiency (DE) (%)	80.29	3.26

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

Cut off Points	Items (N=88)	Interpretation	Action
Difficulty Index (DIF I)			
≤30	6	Difficult	Revise/Discard
31-40	8	Good	Store
41-60	38	Excellent	Store
≥61	36	Easy	Revise/Discard
Discrimination Index (DI)			
<0.15	6	Poor	Revise/Discard
0.15-0.24	4	Good	Store
≥0.25	78	Excellent	Store

Table 3: Distractor analysis (N=264)

Parameter	
Number of items	88
Total distractors Functional distractors	264 212 (80.3%)
Nonfunctional distractors (NFDs)	52 (19.7%)
Items with NFDs (DE between 0-66.6%)	31 (35.2%)
Items with 3 NFDs (DE=0%)	5 (5.7%)
Items with 2 NFDs (DE=33.3%)	11 (12.5%)
Items with 1 NFD (DE=66.6%)	15 (17%)
Items with 0 NFD (DE=100%)	57 (64.8%)
Overall mean DE (Mean ± SD)	80.29±3.26

Table 4. Items (N=31) with nonfunctional distractors and their relationship with DIF I and DI

DIF I (%)	Items with NFDs	DI	Items with NFDs
≤30	1	<0.15	6
31-40	1	0.15-0.24	4
41-60	6	≥0.25	21
≥61	23		
Mean DIF I ± SD (%)	75.30±3.49	Mean DI ± SD	0.43±0.03

When viewed in relation of difficulty level of questions, mean DE was higher (94.43%) in six difficult items than 59.23% in 36 easy items and the mean DE showed statistically significant variation amongst items with changing DIF I (t=2.36, df=40, p=0.02). Mean DE was 88.45% in 78 items with excellent DI compared to 5.55% in six items with poor DI; difference in DE among changing DI was statistically significant with t=9.61, df=82, p<0.0001 (Table 5).

Table 5. Distractor efficiency (DE) of items (N=88) with different values of DIF I and DI

Grading	Difficulty index (DIF I)		Discrimination index (DI)	
	Difficult (≤ 30)	Easy (≥61)	Poor (<0.15)	Excellent (≥0.25)
Number of items	6	36	6	78
DE (%) Mean ± SD	94.43 ± 5.56	59.23 ± 5.95	5.55 ± 5.55	88.45 ± 2.34
Unpaired *t*-test	*t*=2.36, df=40, p=0.02*		*t*=9.61, df=82, p<0.0001*	

DISCUSSION

The most vital aspects of any professional curriculum is the assessment of performance of students at the end semester in the form of internal exams. Their performance provides a feedback about the pace, cognition and content of the course to the teaching faculty [9]. Therefore, the methods which we are using for evaluation of students should be capable of judging the learning objectives, core abilities and skills set as a goal by the concerned faculty. The introduction of MCQs in these exams brought an objectivity in scoring and evaluation. This also enhances the efficacy of teachers to teach a large group of students. The MCQs can be reusable and can be subsequently upgraded and stored in the form of question banks for the repeated use in various combinations and sets. MCQs are mostly used to evaluate the lower order cognition such as recall and also to check student's broad knowledge of the curriculum and learning objectives. If properly framed, these can become equally effective to test the learning objectives of high cognitive level also [9].

There is risk of MCQs being solved correctly by guess work if framed poorly. To consider and improve the quality of MCQs, it is crucial to validate the MCQ questions for the discriminative value of items and to review and authorize the MCQ use. This can be done by developing questions and question banks jointly by the team of evaluators. By continuously developing the MCQ bank, we can help the bench marking processes and also institute the assessment standards which leads to long term effects in promising quality of education.

Each MCQ while being considered for the assessment of students must be studied carefully with the item parameters such as DIF I, DI, and DE to avoid false assessment. DIF I denotes the percentage of students who solved the MCQ correctly and ranges from 0% to 100% [4,10]. DIF I is a misleading term as greater the value of DIF I signifies that the MCQ is easier and vice versa. Hence, some authors called DIF I as ease index also [2]. In the present study, the mean DIF I was 58.2 ± 2.2% which is in the acceptable range (31-60%). Other studies have reported this range as 39–66% [4,6,9,11-13]. Too easy items (DIF I > 60%) will lead to inflated scores, while the difficult items (DIF I ≤ 30%) will result into inflated scores [4]. It is advised that the too easy MCQs should be kept either at the start of the test as ‘warm-up’ questions or can be removed completely, similarly difficult items should be revised for likely areas of disputes, mystifying language or even an incorrect key [2].

DI of an MCQ indicates its ability to distinguish between students of lower and higher abilities and ranges from 0 to 1 [2,10]. It is evident that a question which is either too easy (answered correctly by each student) or too difficult (solved incorrectly by each student) will have nil to poor DI. In the present study the mean DI was 0.69 ± 0.03 within the acceptable range (≥0.25). It was so because 78/88 items had DI more than 0.25. As stated prior, in such cases, students of higher ability solve the questions correctly than those with lower ability. Other studies have reported this range as 0.32-0.37 [6,9,11-13]. Some studies have reported the negative DI too [4,14]. The reasons for negative DI can be incorrect key, confusing framing of question or overall poor preparation of students. MCQs with negative DI are not only of no use, but actually help to fall the validity of the test. According to the cut off points for “good to excellent” of DIF I and DI, there were 46 items as perfect related to 32 and 15 (out of 50) in previous studies [2,4].

The distractors (incorrect alternatives) analysis is performed to evaluate their virtual utility in each MCQ. Items must be revised if students constantly fail to select certain distractors. Such distractors are probably unbelievable and therefore of little use as traps [4]. Therefore, designing of reasonable distractors and decreasing the NFDs is essential part for setting the quality MCQs [15]. MCQ with more functioning distractors decreases DIF I (makes item difficult) and increases DE, conversely more NFD in an item increases DIF I (makes item easy) and reduces DE. Higher the DE more difficult the question and vice versa, which eventually depends on the presence or absence of NFDs in a MCQ. Mean DE in present study was 80.3 ± 3.3% lower than DE reported elsewhere in a similar type of studies [4,6,12].

Item analysis is an important process performed after the examination which provides information regarding the reliability and validity of an item or test [16]. It aids in detecting specific technical flaws and thus provides information for improving test item. Thus, we conclude that the analysis of items strengthen the future question bank. Also, discussion of the analysis result with the faculties helps in modification of teaching methodology and outcome of learning. Therefore, item analysis is a simple and feasible method of assessing valid MCQs in order to achieve the ultimate goal of medical education.

REFERENCES

[1] M Chandratilake, M Davis, G Ponnamperuma. *Intern J Med Edu*. **2009,** 1,1-7.

[2] MR Hingorjo, F Jaleel. *J Pak Med Assoc*. **2012,** 62,142-146.

[3] TM Haladyna, SM Downing, MC Rodriguez. *Appl Measurement Edu*. **2002,** 15(3), 309-334.

[4] S Gajjar, R Sharma, P Kumar, M Rana. *Indian J Comm Med*. **2014,** 39(1), 17-20.

[5] MR Sharif, SM Rahimi, M Rajabi, M Sayyah. *J Sci Tech*. **2014,** 4(10), 565-569.

[6] V Kolte. *Indian J Basic Appl Med Res*. **2015,** 4(4), 320-6.

[7] T Singh, P Gupta, D Singh. Test and Item analysis. In: Principles of Medical education.3rd ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd:**2009,** 70-77.

[8] J Considine, M Botti, S Thomas. *Collegian*. **2005,**12,19-24.

[9] T Kumar, Y Kumar, K Jha, R Singh. *Ind J Physiol Pharmacol*. 2017; 61(1):70-75.

[10] M Tarrant, J Ware, AM Mohammed. *BMC Med Educ*. **2009,** 9,1-8.

[11] SK Namdeo, S Sahoo. *Int J Res Med Sci*. **2016,** 4, 1716-1719.

[12] AR Menon, PN Kannambra. *Nat J Lab Med*. **2017,** 6(2), MO07-MO10.

[13] G Mehta, V Mokhasi. *Int J Health Sci Res*. **2014,** 4(7), 197-202.

[14] www.freepdfdb.com/pdf/item-analysis-scantron

[15] MC Rodriguez. *Educ Measure*. **2005,** 3-13.

[16] MJC Botti, S Thomas. *Collegian*. **2005**, 12, 19-24.