RESEARCH PAPER | Medical Science | Volume : 6 | Issue : 3 | March 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50

Item Analysis to Assess Quality of MCQs.

KEYWORDS: Difficulty Index, Discrimination index, Distractor Efficiency, MCQs as item.

Dr. Swapnagandha S. Halikar
Associate Professor, Dept. of Ophthalmology, Smt. Kashibai Navale Medical College, Pune.

Dr.Veerendra Godbole
Associate Professor, Dept. of Ophthalmology, Smt. Kashibai Navale Medical College, Pune.

Dr.Saurabh Chaudhari
PG student, Dept. of Ophthalmology, Smt. Kashibai Navale Medical College, Pune.

ABSTRACT *Introduction: Assessment of medical undergraduate students give insight about their knowledge and competencies. MCQs (items) are being increasingly used in such assessments. Selection of quality MCQs is necessary to assess the knowledge of the students & to differentiate the students of different abilities in correct manner. This study was conducted at Smt. Kashibai Navale Medical College to evaluate MCQs and to develop valid pool of MCQs*

Materials and methods: 60 students of VII semester MBBS (Ophthalmology) appeared for an internal assessment in June 2015. 20 items were analyzed for Difficulty index, Discrimination index and Distractor efficiency.

Results: After statistical analysis of data, it was observed that Difficulty index i.e. 'p' value of analyzed MCQs ranged from 5% (lowest) to 95% (Highest) & Discriminative index i.e. 'd' value ranged from - 0.2 (lowest) to 0.65 (Highest). Total 35% items were in acceptable range of difficulty level ('p' value 30 – 70%), 50% items were easy & 15 % items were difficult. Discrimination index of 30% items was excellent (d value>0.35), while 15% items were having poor discriminative power. Distractor Efficiency (DE) of 20% items was 66.6% whereas 50% items had 0% DE.

Conclusion: Item Analysis is simple procedure for evaluation of validity & reliability of MCQs thus improving its quality thereby strengthing the question bank.

Introduction:
Assessment of medical undergraduate students gives insight about their knowledge and competencies. MCQs are being increasingly used in such assessments. Selection of quality MCQs is necessary to assess the knowledge of the students & to differentiate the students of different abilities in correct manner.MCQs are used as reliable tool to evaluate learning performance of the students and it can assess cognitive, affective as well as psychomotor domain[1, 2]. Item analysis enables identifying good MCQs based on difficulty index (DIF I) also denoted by FV (facility value) or *P*-value, discrimination index (DI), and distractor efficiency (DE).[3,4]It is a valuable yet relatively simple procedure performed after the examination that provides information regarding the reliability and validity of a test[5]. Item analysis examines student's responses to individual test item, assesses the quality of those items & helps to improve/revise the items & the test.

Materials & Methods:
60 students of VII semester MBBS (Ophthalmology) appeared for an internal assessment in June 2015. The examination consisted of MCQs, SAQs & LAQs. There were 20 'single response type' MCQs. Half an hour was allotted to solve 20 MCQs and 0.5 marks were allotted for each MCQ. There was no negative marking for wrong answers. All MCQs had single stem with four options including one correct answer (key) & other three were incorrect alternatives (distractors). Each item was analyzed for Difficulty Index (Dif I), Discrimination Index (DI) and Distractor Efficiency (DE). Data obtained was entered in MS Excel & analyzed. Total score of 60 students was arranged in descending order & total numbers of students were divided into three groups. One group consisting of 1/3 of total students with higher marks and labeled as students of higher ability (H) & 2nd group consisting of 1/3 of total students with low marks labeled as students of low ability (L). Middle 1/3 (20 students) were not considered in the study. Total 20 MCQs with 60 distractor were analyzed & based on this data various indices like difficulty index discrimination index and distractor efficiency were calculated with the following formulae[1, 6, 7].

Difficulty index (DIF I) = X100
Here: N = total no of students in high & low group
H= no. of students answering correctly in high group
L= no. of students answering correctly in low group

Difficulty index
30-70 = Acceptable
< 30 = Difficult
>70 = Easy

DIF I describes the percentage of students who answered the item correctly and ranges between 0 and 100%[8]. DIF I is a misnomer as bigger is the value of DIF I, easier is the item and vice versa; hence, it is also called as ease index[9].

Discrimination index (DI) =

Discrimination index
0.35 and over: Excellent question
0.20 to 0.34 : Good question
>0 to 0.20 : Acceptable question
0 : Cannot discriminate
<0 :Poor question

Here: N = total no of students in high & low group
H= no. of students answering correctly in high group
L= no. of students answering correctly in low group

Volume : 6 | Issue : 3 | March 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50

DI is the ability of an item to differentiate between students of higher and lower abilities and ranges between 0 and 1. Higher the value of DI, item is more 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[10]. There are instances when the value of DI can be <0 (negative DI) which simply means that the students of lower ability answer more correctly than those with higher ability. Such situations though undesirable, happen due to complex nature of item making it possible for students of lower ability to select correct response without any real understanding. Here a student of lower ability by guess, selects correct response; while a good student suspicious of an easy question, takes harder path to solve and end up to be less successful[4].

An item contains a stem and four options including one correct (key) and three incorrect (distractor) alternatives[4, 8]. Here, non functional distractor (NFD) in an item is option (s) (other than key) selected by <5% of students; alternatively functional or effective distractors are those selected by 5% or more participants [9, 10]. DE is determined for each item on the basis of the number of NFDs in it and ranges from 0 to 100%. If an item contains three, two, one or nil NFDs then DE will be 0, 33.3, 66.6, and 100%, respectively[9].

Results:

Table1:

Analysis of items according to Difficulty Index (p value)

Difficulty Index 'p' value	Analysis of item	Total number of items
<30%	Difficult	3 (15%)
30-70%	Acceptable	7 (35%)
>70%	Easy	10 (50%)

Table 2:

Analysis of item according to Discrimination Index (d value)

Discrimination Index (d)	Discrimination Power	Total number of items
≥0.35	Excellent	6 (30%)
0.2 to 0.34	Good	4 (20%)
>0 to 0.2	Acceptable	6 (30%)
0	Can't discriminate	1 (5%)
<0	Poor	3 (15%)

Table 3:

Analysis of items according to Distractor Efficiency

Items with 0 NFD DE=100%	Items with 1 NFD DE=66.6%	Items with 2 NFD DE=33.3%	Items with 3 NFD DE=0%
0 (0%)	4 (20%)	6 (30%)	10 (50%)

NFD= Non-functional distractor

Discussion:

Various methods of assessments are used for evaluation of medical students, one of which is multiple choice questions (item) .MCQs can assess cognitive, affective as well as psychomotor domain of the students. It is preferred methods over others because of objectivity, elimination of assessor bias and wide coverage of the subject. Item analysis is a valuable procedure in providing information regarding reliability & validity of an item by calculating difficulty index, discrimination index and distractor efficiency[3, 4]. This is because if the MCQ is flawed then it itself becomes distracting and assessment can be false, thus helps in differentiating good MCQs from bad MCQs. It also helps to revise / store or discard & to develop pool of valid MCQs[3, 10, 11].

In the present study, after analysis of data, it was observed that DIF I (p value) of MCQs ranged from 5% (L) – 95% (H).The mean of DIF I in this study was 61.75±28.20. DIF I of 7(35%) items was in acceptable range (30 – 70%). Others had proposed this range as 41 -60%[11]. 10(50%) were easy items with DIF I value of >70% and only 3(15%) items were difficult with p value of <30% which corresponded with other studies where mean of DIF I was 39.4 ± 21.4 [12], 52.53 ± 20.59[6] One study showed that 62% items had p value (30-70%), 23 % were too easy (p >70%) and 15% were too difficult (p<30%) [6].

On reviewing easy items, it was found that they were from 'must know ' portion but the distractors given in these items were mostly nonfunctional distractors, so these items were revised, nonfunctional distractors were replaced by functional distractors and then added to question bank. It is advisable to place easy items at the beginning of the paper so as to boost the confidence of overall students and taken as warm up questions. Difficult items were used to select toppers.

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.

Mean DI in present study was 61.75±28.20. Total 6(30%) items had excellent discrimination power (≥0.35). Items with good discriminative power were 4(20%) while 6(30%) items were in acceptable range. Discriminative power of 1(5%) item was non-discriminative (DE=0) while 3(15%) items were having poor discriminative power or negative discriminative power. These findings corresponded with earlier studies which revealed 40% items with DI >0.35, 42% with DI between 0.2 and 0.34 and 18% with DI < 0.20[7]. Another study showed 29% items with DI>0.4%, 46% items with DI between 0.2-0.39 and 21 % items with DI < 0.19[13]. Si - Mui et al showed that items with good discriminating potential tend to be moderately difficult items[14]. The relationship between Dif I and DI is not linear, but predicted as dome shaped[6, 15]. For easy items, discrimination may be poor. This is because both, high and low achievers can answer correctly. Items with very high discrimination index are given in Entrance exams, in which main purpose of test is to select only few good students from many. Items with negative DI are not only useless, but actually serve to decrease the validity of the test[4] and should be removed from question bank. Some studies have shown negative DI in 20% [12] and 4% [9] items. In present study, 3(15%) items were having poor discriminative power or negative discriminative power. These 3 items were having more nonfunctional distractors which were replaced by functional distractors and added to question bank.

Most difficult task in formatting good quality MCQs is writing appropriate options to the correct answer. A distractor analysis gives an opportunity to study the responses made by students on each alternative of the item. NFDs should be removed from the item or be replaced with a more plausible option[16]. Analysis of the distractors (incor-

Volume : 6 | Issue : 3 | March 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50

rect 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[4]. Therefore, designing of plausible distractors and reducing the NFDs is important aspect for framing quality MCQs[15]. 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, there were total 60 distractors, out of which 27(45%) were non-functional distractors and 14(23.33%) were functional distractors and 19(31.66%) distractors had nil response. On the basis of number of NFDs, items with DE 66.6% were 4(20%), items with DE 33.3% were 6(30%) and items with DE 0 were 10(50%). No items were having DE of 100%. Mehta G et al have shown, in their study, with fifty MCQs, having150 distractors, 53(35.33%) were found to be NFDs, 28(18.66%) were functional distractors and 69(46.01%) distractors had nil response. The number of items having NFDs was found to be 33(66%). On the basis of number of NFDs, items with DE 66.6% were 18(54.4%), items with DE 33.3% were 9(27.27%) and items with DE as 0 were 6(18.18%). The remaining 17 items with three functional distractors had DE as 100%[17]. Gajjar et al have shown that, in a total of 150 distractors, 133(89.6%) were functional distractors, and 17(11.4%) were NFDs. Items with NFDs were 15(30%) out of which 13items had DE of 66.6% and 2 items had DE of 33.33%[12]. Students' performance depends on how distractors are designed. Analysis of the distractors, identifies their errors, so that they may be revised, replaced or removed[18].

Conclusion:

Item Analysis is simple procedure for evaluation of validity & reliability of MCQs. Item analysis and storage of MCQs with their indices provides opportunity for an examiner to select MCQs of appropriate difficulty level as per the need of assessment and decide their placement in the question paper. MCQs are preferred method of objective assessment of the students as it covers wide area of the subject in a short period of time. Therefore selection of good MCQs can judge knowledge of the students.

REFERENCES

1. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy(Online) 2011(Cited 2011 Feb2). Available from URLhttp://web.uct.ac.za/projects/cbe/ mcqman/mcqappc. html.
2. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. Tech Learn med 1993;5:107-15
3. Scorepak®: Item analysis. [Last citied on 2013 Apr 13]. Available from:www.washington.edu./oea/score1/htm
4. 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 .
5. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian 2005;12:19-24
6. Pande SS et al. Correlation between difficulty and discrimination indices of MCQs in formative exam in Physiology. South East Asian Journal of MedicalEducation 2013;7:45-50
7. Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam.International Journal of Scientific Research 2013;2:279-280
8. 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 .
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. [PubMed]
10. 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.
11. Ahmadabad: Regional Training Centre Smt. NHL Medical College; 2012. Manual of Basic Workshop in medical education technologies.
12. GajjarS, Sharma R, Kumar P,Rana M. Item and test analysis to identify quality Multiple Choice Questions (MCQs) from an assessment of medical students of Ahemdabad, Gujarat. Indian journal of Community Medicine 2014;39:17-20
13. Carroll RG. Evaluation of Vignette-type examination items for testing medical physiology. Am J Physiol 1993;264:S11-5
14. Si – Mui S, Isaiah R . Relationship between item difficulty and discrimination indices in true/false – type multiple choice questions of a para – clinical multidisciplinary paper.Ann Acad Med 2006;35:67 – 71.
15. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four response type multiple choice questions of pharmacology summative tests of year II MBBS students. Ie JSME2013;6:41- 46
16. Haladyna TM, Downing SM. Validity of a taxonomy of multiple choice item- writing rules. ApplMeasEduc 1989;2:51-78
17. Mehta G, Mokashi V. Item analysis of Multiple Choice Questions-An assessment of the assessment tool. International Journal of Health science & Research 2014;4:197-201
18. Dufresne RJ, Leonard WJ, Gerace WJ. Making sense of students answers to multiple choice questions. Phys Teach 2002;40:174-180