Item Analysis: An assessment of the Assessor's Items

Education

Dr. Reetika Sharma	Assistant Professor, Department of pathology, Pondicherry Institute of Medical Sciences, Pondicherry
Dr. Renu G'Boy Varghese	Professor, Department of pathology, Pondicherry Institute of Medical Sciences, Pondicherry
Kandasamy Ravichandran	Biostatistician, Pondicherry Institute of Medical Sciences, Pondicherry

ABSTRACT

AIM: To assess the quality of multiple choice questions and take further intervention **MATERIALS AND METHODS:** Total of 30 multiple choice questions were taken to find out difficulty index (DIF I) , discrimination index (DI) , and distractor efficiency (DE) . **RESULTS :** 21 items showed DIF I between 30-70%, 9 items had   >70% . Out of 30 items, 21 had DI index between 0.2 – 0.35 while 9 had > 0.35. None of the items showed DI < 0.15. Out of total 30 items, 18 (60%) had nil NFD {DE= 100%} , 6 (20%) had 1NFD { DE= 66.6%}, 4 (13.3%) had 2 { DE= 33.3%} and 2 (6.7%) had 3 NFDs { DE=6.3%} . Thus mean ± SD value for DIF I came out as 58.84 ± 19.16 % and for DI and DE came as 0.31 ±   0.14 and DE = 77.76 ± 31.98 respectively.

CONCLUSION : This study emphasize that questions having average difficulty and high discriminating power with presence of functional distractors should be incorporated into further tests. Very difficult and very easy items need to be properly reconstructed.

KEYWORDS:

Item analysis, difficulty index, discrimination index, and distractor efficiency

INTRODUCTION

Evaluation is an essential part of teaching-learning curriculum. Today Multiple Choice Questions (MCQs) is the most frequently used tool for assessing the knowledge of students. MCQs or items are used in assessments of learning and competencies. MCQs based evaluation in addition to assessing knowledge also evaluates understanding and analyzing power of students.

Designing good MCQs is a multifaceted, difficult and lengthy process. "Item analysis" examines student responses to individual test items (MCQs) to assess the quality of items and test as a whole to improve items and the test. Thus item analysis assesses the assessment tool i.e. MCQs for the benefit of both student and teacher. 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).[1] The aim of this study to assess the quality of multiple choice questions, for creating a viable question bank for future use and to identify low achievers to make improvement in them for better results

MATERIALS AND METHODS

This study was conducted   in the Department of Pathology in a medical college in south India in 2015. Every month an internal examination is conducted in pathology for 3rd semester students which include one essay, five short notes and five MCQs. In this study we took six internal question papers containing five MCQs each. Thus total of 30 MCQs were considered for analysis. Each MCQ had one best answer and three distractors. There was no negative marking. Data obtained was entered in MS Excel 2007 and analyzed. Those students who attended all internals (six) were included in the study and those who were absent in any one of these were excluded from the study. Thus out of 96 students 76 were included in the study. The scores of all the students were arranged in order of merit. The upper one third students were considered as high achievers and lower third as low achievers.

Each item/ MCQ was analysed for:

Difficulty Index (DIF I) or Facility value or p value= [(H+L) / N] x 100

Discrimination index (DI) or d value = (H-L) x 2 / N

Distractor Effectiveness (DE) and non-functional distractor (NFD) were also calculated. 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. DE is determined for each item on the basis of the number of NFDs in it and

ranges from 0 to 100%.

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

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

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

RESULTS

Total 30 MCQs and 90 distractors were analyzed. Score ranges from 9 to 27 (out of 30).

Table 1: Score wise distribution of the students (n=76)

Score	Group	Number (percentage)
9 – 15	Low	24 (31.6)
16 – 19	Middle	27 (35.5)
20 – 27	High	25 (32.9)

Table 2: Difficulty Index

Cut off points (p value)	Items (N =30)	Interpretation
30-70 %	21 (70%)	Acceptable
>70 %	9 (30%)	Too easy
<30 %	0	Too difficult
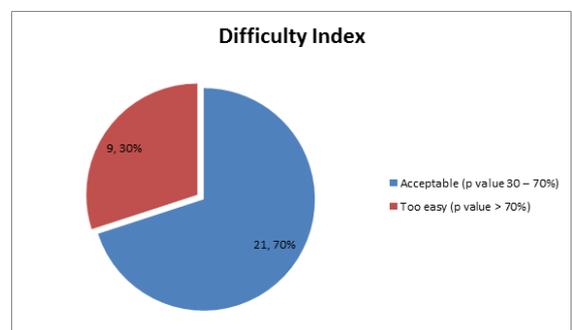
Fig. 1

Table 3: Discrimination Index

Cut off points	Items (N =30)	Interpretation
0.2 - 0.35	21 (70%)	Good
> 0.35	9 (30%)	Excellent
< 0.15	0	Poor

Discrimination Index

9, 30%

21, 70%

■ Good (d value 0.20 - 0.35)
■ Excellent (d value > 0.35)
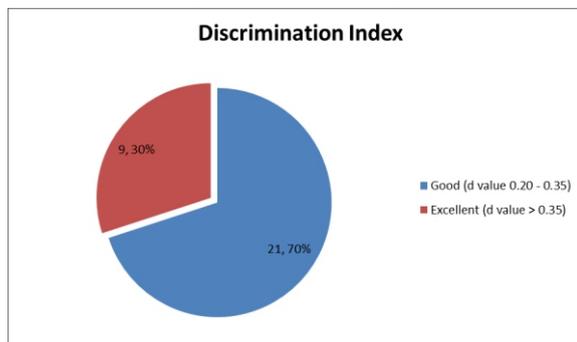
Fig 2

Table 4: Distractor Effectiveness

	Frequency	Percentage (%)
Items with 0 NFD	18	60.0
Items with 1 NFD	6	20.0
Items with 2 NFD	4	13.3
Items with 3 NFD	2	6.7
Total	30	100.0

Table 5: Comparison of Dif I, DI and DE of the MCQs (items)

Parameter	Dif I	DI	DE
Range	30.61 - 93.88%	0.16 - 0.73	0-100%
Mean ± SD	58.84 ± 19.16	0.31 ± 0.14	77.76 ± 31.98

Thus out of 76 students,   25 were in high achiever group and 24 were in low achiever group. 21 items showed Dif I between 30-70%, 9 items showed >70% while none of the items had Dif I <30% . Out of 30 items, 21 had DI index between 0.2 – 0.35 while 9 had > 0.35. None of the items showed DI < 0.15. Out of total 30 items, 18 (60%) had nil NFD {DE= 100%} , 6 (20%) had 1NFD { DE= 66.6%}, 4 (13.3%) had 2 { DE= 33.3%} and 2 (6.7%) had 3 NFDs { DE=6.3%} .

Thus mean ± SD value for Dif I came out as 58.84 ± 19.16 % and for DI and DE came as 0.31 ±   0.14 and DE = 77.76 ± 31.98 respectively.

DISCUSSION

Quality medical care depends upon the development of knowledgeable, skilled, and competent medical personnel. The effective measurement of knowledge is an important component of both medical education and practice. [2] It is important for us to evaluate our MCQs/ items to see how effective they are in assessing the knowledge of our medical students. The analysis of the MCQs helps to assess the quality of individual test items and test as a whole. It also helps to identify the contents in subject which lacks understanding and need greater emphasis and clarity, so that we can improve or change the methodology of teaching.

Each item (MCQ) used in assessment must be evaluated based on DIF I, DI, and DE because if an item is faulty then the assessment can be false. Difficulty index is the proportion of total students in the two groups who have answered the item correctly. [3] Items with a p value between 30 – 70% are considered as acceptable while 50-60% is ideal. [3] Too difficult items (DIF I ≤ 30%) will lead to low scores, while the easy items (DIF I > 60%) will result in high scores and a decline in motivation.[4] Too easy items can be placed either at the start of the test so that students don't loose interest from the test or should be removed altogether.

In our study, mean of Dif I was 58.84 ± 19.16. Karelia BN et al [5], showed a range of mean ± SD between 47.17±19.77 to 58.08±19.33. In their study they found that 61% items were in acceptable range (p 30-70%), 24 % items were too easy (p>70%) and 15 % items were too difficult (p< 30%) while in our study 70% items were in acceptable range and 30% items were too easy. Another study done by Mehta G et al [3] found that p value of 31 (62%) items was 30-70%, 16(32%) items >70% and 3(6%) items <30%.

DI of an item indicates its ability to differentiate between students of higher and lower abilities. This means that DI will be poor to nil for too difficult (done wrongly by everyone) or too easy (attempted correctly by everyone) questions. Value of DI normally ranges between 0 and 1 while acceptable cut off point is taken as 0.15. [1] Mean DI in the present study is 0.31 ± 0.14 which is within the normal range. Items with DI > 0.35 were 9 (30%) only and DI between 0.2 - 0.35 were 21 (70%). Dif I (p value) and DI both are reciprocal to each other. It means that higher the Dif I (p value) less difficult will be that item and lower will be its DI. Thus questions with low p value are considered as good discriminator. Sometimes DI value can be negative. It means that students with lower abilities have answered correctly. This is due to complex nature of question which is answered by these students by guess only. [1, 6] One study done by Patel KA et al [7] found DI > 0.35 in 40% and other study it was 29% items with DI > 0.40. [6]

Writing appropriate options while framing a question is one of the difficulties faced by the examiner. A distractor analysis calculate the responses made by students on each alternative of the item.[3]

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.[1] Higher the DIF I , more easy the question will be and which in turn depend upon the presence of NFD. NFDs should be removed from the item or be replaced with a more reasonable option. [3] In present study, out total 30 items, 18 (60%) had nil NFD {DE= 100%} , 6 (20%) had 1NFD { DE= 66.6%}, 4 (13.3%) had 2 { DE= 33.3%} and 2 had 3 NFDs { DE=6.3%} . Thus 18 (60%) items had FD and 12(40%) items had NFD The mean DE came out as 77.76 ± 31.98 while 88.6 ± 18.6% .[1]

In   the study done by Mehta et al [3] with fifty MCQs, having 150 distractors, 53(35.33%) were found to be NFDs, 28(18.66%) were FD and 69(46.01%) distractors had nil response. In the present study however, we didn't find any nil response.

CONCLUSION

Item analysis is a simple procedure regarding the reliability and validity of an item/test by calculating DIF I, DI, and DE. An ideal item (MCQ) will be the one which has average difficulty (DIF I between 31 and 60%), high discrimination (DI ≥ 0.25) and maximum DE (100%). This study emphasize that questions having average difficulty and high discriminating power with presence of functional distractors should be incorporated into further tests. Very difficult and very easy items need to be properly reconstructed. In this study, the majority of items (21, 70%) fulfilled the criteria of acceptable difficulty and good discrimination, which means the MCQs selected were of good quality. However too easy items (9, 30%) , will be reconstructed / or eliminated from future tests.

References:
1.   Gajjar S, 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.
2.   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.
3.   Mehta G, Mokhasi V. Item analysis of multiple choice questions- an assessment of the assessment tool. Int J Health Sci Res. 2014;4:197-202.
4.   Eaves S, Erford B. The Gale group: The purpose of item analysis, item difficulty, discrimination index, characteristic curve. Available from: www.education.com/ reference/article/itemanalysis/ [Last cited on 2013 Apr 15].
5.   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.
6.   Carroll RG. Evaluation of Vignettetype examination items for testing medical physiology. Am J Physiol 1993;264:S11-5.
7.   Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam. International Journal of Scientific Research 2013;2:279-28.