International Journal of Health Sciences and Research
www.ijhsr.org ISSN: 2249-9571

Original Research Article

Item Analysis of Multiple Choice Questions- An Assessment of the Assessment Tool

Gyata Mehta[1], Varsha Mokhasi[2]

[1]Assistant Professor, [2]Professor and Head,
Department of Anatomy, Vydehi Institute of Medical Sciences and Research Centre, Bangalore- 560066

Corresponding Author: Gyata Mehta

Received: 19/05//2014 *Revised: 09/06/2014* *Accepted: 13/06/2014*

ABSTRACT

Objectives- The aim of the study was to assess the quality of multiple choice questions, for creating a viable question bank for future use. The purpose was also to identify the low achievers and their learning difficulties which can be corrected by counselling or modifying learning methods.
Methods-The study was conducted in the department of Anatomy. A hundred First-year MBBS students took the MCQs test comprising of fifty questions. There was no negative marking and evaluation was done out of fifty marks and 50% score was the passing mark. Post validation of the paper was done by item analysis. Each item was analysed for Difficulty index, Discrimination index and Distractor effectiveness.
Results- Difficulty index of 31(62%) items was in the acceptable range (p value 30-70%), 16(32%) items were too easy (p value >70%) and 3(6%) items were too difficult(p value <30%). Discrimination index of 26 (52%) items was excellent (d value>0.35), 9(18%) items was good (d value 0.20-0.34) and 15(30%) items were poor (d value<0.2%). A total of fifty items had 150 distractors. Amongst these, 53(35.3%) were nonfunctional distractors, 38(18.6%) were functional distractors and 69(46.06%) had nil response i.e. not attempted by any student. On the basis of non-functional distractors, distractor effectiveness of each item was assessed. Inter- relationship between these indices was analysed.
Conclusion - This study inferred that items having average difficulty and high discriminating power with functional distractors should be incorporated into future tests to improve the test development and review.

Key words- item analysis, multiple choice questions, difficulty index, discrimination index, distractor effectiveness

INTRODUCTION

Evaluation is an important component of a teaching-learning curriculum. A significant application of evaluation is for continued monitoring of learning activities for giving a feedback to students and teachers. Today Multiple Choice Questions (MCQs) is the most commonly used tool for assessing the knowledge capabilities of medical students.

However it is said that MCQs emphasize recall of factual information rather than conceptual understanding and interpretation of concepts. [1] There is more to writing good MCQs than writing good questions. Properly constructed MCQs can assess higher cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, instead of just testing recall of isolated facts. [2,3]

Designing good MCQs is a complex, challenging and time consuming process. Having constructed and assessed, MCQs need to be tested for the standard or quality. Item analysis examines the student responses to individual test items (MCQs) to assess the quality of those items and test as a whole. [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] Thus item analysis assesses the assessment tool for the benefit of both student and teacher.

We took this study to analyze the quality of MCQs, to improve the items that needed modification, for creating a viable question bank for subsequent use. The purpose was also to identify the low achievers and their learning difficulties which can be corrected by counselling or modifying learning methods. The teachers would also get a feedback on the efficacy of their teaching, for improvement of teaching skills in the future.

MATERIALS AND METHODS

The study was conducted in the department of Anatomy as a 'part end' assessment. A hundred First-year MBBS students took the MCQs test comprising of fifty questions of type 'A' with single best response. There was no negative marking and the time allotted was one hour. Pre-validation of the paper was done by scrutinization by the Head of Department. Evaluation was done out of fifty marks and 50% score was the passing mark.

Post validation of the paper was done by item analysis. 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 was analysed for:

I Difficulty Index (Dif I) or Facility value or p value using the formula $p = H + L / N \times 100$

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-responders)

II Discrimination index (DI) or d value using the formula $d= H-L\times 2/N$

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

III Distractor Effectiveness (DE) or Functionality

Interpretation

Difficulty index is merely the proportion of total students in the two groups who have answered the item correctly. In general, items with a p value between 30 – 70% are considered as acceptable. Amongst these, items with p value between 50-60% are ideal. Items with p value less than 30% (too difficult) and more than 70% (too easy) are not acceptable and need modification.

The Discrimination index, also called point biserial correlation is a measure of the item to discriminate between students of higher and lower abilities and ranges between 0 and 1. In general the 'd' value between 0.20 and 0.35 is considered as good. Items with DI more than 0.35 are considered as excellent and those with DI less than 0.20 are considered as poor.

An item contains a stem and four options including one correct (key) and three incorrect (distractor) alternatives. Non Functional Distractor (NFD) in an item is the option, other than the key selected by less than 5% of students and functional or effective distractor 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. [6]

RESULTS

I Difficulty Index

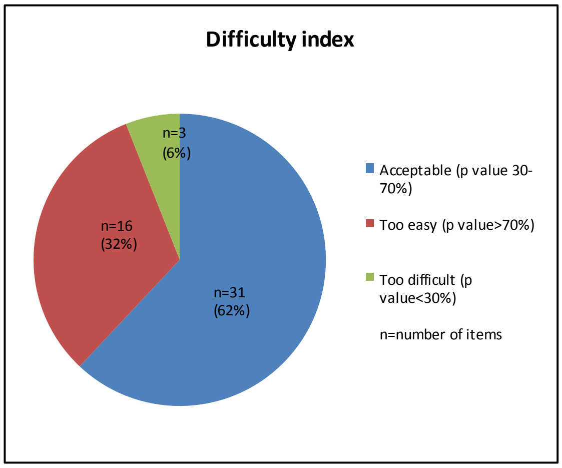

Figure 1- Difficulty index (p value) of each MCQ item.

The number of items having ideal Dif I (p value 50-60%) was 12(24%).

II Discrimination Index

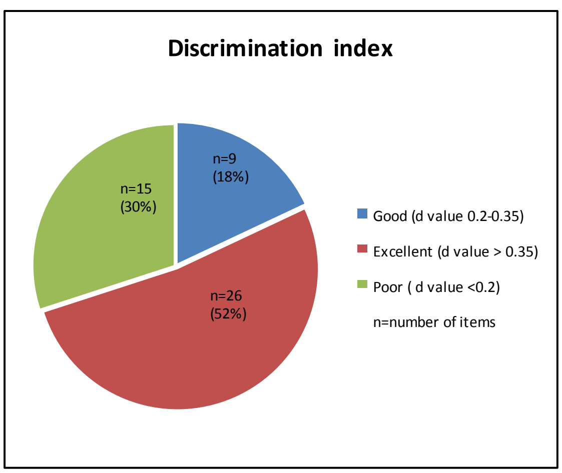

Figure 2- Discrimination index of each MCQ item.

III Distractor Effectiveness

Table 1 - NFDs and Distractor Effectiveness (DE) of each MCQ item.

Total number of items 50			
Items with 0 NFDs DE = 100%	Items with 1 NFD DE= 66.6%	Items with 2 NFDs DE= 33.33%	Items with 3 NFDs DE= 0%
17 (34%)	18(54.54%)	9 (27.27%)	6(18.18%)

NFD-Non Functional Distractor, DE- Distractor Effectiveness

Table 2- Comparison of Dif I, DI and DE of the MCQ items.

Parameter	Dif I	DI	DE
Range	16.66 - 96.96%	0- 0.66	0-100%
Mean ±SD	63.06 ± 18.95	0.33 ± 0.18	63.97 ± 33.56

Dif I- Difficulty Index, DI- Discrimination Index, DE- Distractor Effectiveness

A total of fifty items had 150 distractors. Amongst these, 53(35.3%) were NFDs, 38(18.6%) were functional distractors and 69(46.06%) had nil response i.e. not attempted by any student.
A total of 12 (24%) items were found to be 'ideal' having Dif I (p value 50-60%) and DI > 0.35.
Amongst these, 5 items had DE 100% (0 NFD), 6 items had DE66.6% (1 NFD) and 1 item had DE 33.33% (2NFDs).

DISCUSSION

Post examination analysis of the MCQs helps to assess the quality of individual test items and test as a whole. It also helps to identify the subject content which lacks understanding and need greater emphasis and clarity, by improving or changing the methodology of teaching. Poor items can be modified or removed from the store of questions.

Previous studies have proposed the mean of Dif I as 39.4±21.4%[6], 52.53±20.59. [7] Karelia B, showed a range of mean ±SD between 47.17±19.77 to 58.08±19.33 in a study conducted over a period of five years. [8] They also showed 61% items in acceptable range (p 30-70%), 24 % items (p>70%) and 15 % items (p< 30%). Other studies showed that 62% items had p value (30-70%), 23 % were too easy (p >70%) and 15% were too difficult (p<30%).[7]Patel KA and Mahajan NR showed 80% of items in the acceptable range (p30-70%) and 20% in the unacceptable range (p >70% &<30%). [9] Our findings corresponded with the previous studies having a mean of Dif I as 63.06± 18.95. The p value of 31 (62%) items was in the acceptable range (30-70%), 16(32%) items >70% and 3(6%) items <30%.

Higher the Dif I, lower is the difficulty of the question. The Dif I and DI are often reciprocally related. Questions having high p value (easier questions) discriminate poorly; conversely questions with a low p value are considered to be good discriminators. [10] Value of DI normally ranges between 0 and1. There are instances when the value of DI can be less than 0 (negative DI), which simply means that the students of lower ability answer more correctly than those with higher ability. This is probably due to complex nature of item, making it possible for students of lower ability to select correct response by guess without any real understanding, while a good student suspicious of any easy question, takes a harder path to solve and ends up to be less successful. [6]

In the present study, the mean of DI was 0.33± 0.18. Items with DI > 0.35 were 26(52%), DI between 0.2 and 0.34 were 9(18%) and DI <0.2 were 15(30%). There were no items with negative DI. Some studies have shown negative DI in 20% [6] and 4% [11] MCQ items. Probable explanation was wrong key, ambiguous framing of questions or generalized poor preparation of students. [6] Items with negative DI decrease the validity of the test and should be removed from the collection of questions. Earlier studies have revealed 40% items with DI >0.35, 42% with DI between 0.2 and 0.34 and 18% with DI < 0.20. [9] 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. [10] It has been seen that the relationship between Dif I and DI is not linear, but predicted as dome shaped. [7,8]

A practical difficulty faced by teachers in formatting high 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. [12]

In a study conducted on 514 items and 1542 distractors, 35.1% were NFDs, 52.2% were functional distractors and 10.2% were not chosen by any student. [13] Another review of functioning distractors in 477 items on four MCQ assessments showed 38% items had NFDs and items with three functional distractors ranged from only 1.1 to 8.4%. [14] In the present study with fifty MCQs, having 150 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 MCQ 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%. 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 2items had DE of 33.33%. [6]

Students' performance depends on how distractors are designed. [15] Analysis of the distractors, identifies their errors, so that they may be revised, replaced or removed. Based on the cut off points for 'good to excellent' for Dif I and DI, items were considered as 'ideal' having Dif I (50-60%) and DI>0.35 .Our study profiled 12(24%) items as ideal as compared to 15(30%). [8] Amongst these 12 MCQs, 6 items had DE 66%, 5 items had DE 100% and 1 item had

DE 33.3%. Hence it was proved that having one or two NFDs in an item is considered better than having no NFDs.

CONCLUSION

Item analysis is largely used for creating a viable question bank and to assess class performance as a part of formative assessment. This study inferred that items having average difficulty and high discriminating power with functional distractors should be incorporated into subsequent tests. It is only through the iterative process of item analysis and improvement that pedagogically and psychometrically sound tests can be developed. [13] Item analysis helps tremendously to achieve better teaching, better learning and in the long term better tests. [16]

REFERENCES

1. Srivastava A, Dhar A, Agarwal CS. Why MCQ? Indian Journal of Surgery2004;66:246-248
2. 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.
3. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. Tech Learn med 1993;5:107-15
4. Singh T, Gupta P, Singh D. Test and Item analysis.In :Principles of Medical Education.3rd ed. New Delhi: Jaypee Brothers Medical Publishers(P) Ltd:2009.p,70-77
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. 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
7. Pande SS et al. Correlation between difficulty and discrimination indices of MCQs in formative exam in Physiology. South East Asian Journal of Medical Education 2013;7:45-50
8. 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
9. Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam. International Journal of Scientific Research 2013;2:279-280
10. Carroll RG. Evaluation of Vignette-type examination items for testing medical physiology. Am J Physiol 1993;264:S11-5
11. Hingorjo MR, Laleel F. Analysis of One -Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. J Pak Med Assoc 2012;62:142-147
12. Haladyna TM, Downing SM. Validity of a taxonomy of multiple choice item-writing rules. ApplMeasEduc 1989;2:51-78
13. Tarrant M, WareJ, Mohammed AM. An assessment of functioning and nonfunctioning distractors in multiple choice questions: a

descriptive analysis. BMC Medical Education 2009 ;9:40
14. Haladyna TM, Downing SM. How many options is enough for a multiple choice test item? Educ Psychol Meas 1993;53:999-1010
15. Dufresne RJ, Leonard WJ, Gerace WJ. Making sense of students answers to multiple choice questions. Phys Teach 2002;40:174-180
16. Ananthakrishnan N. Item Analysis validation and banking of MCQs. In: Medical Education Principles andP.2nd ed. Ananthkrishnan N, Sethuraman KR, kumar S. Alumini Association of National Teacher Training Centre, JIPMER, Pondichery, India. p,131-137.

How to cite this article: Mehta G, Mokhasi V. Item analysis of multiple choice questions- an assessment of the assessment tool. Int J Health Sci Res. 2014;4(7):197-202.

****