International Journal of Sciences & Applied Research

www.ijsar.in

Original Article

Improving Multiple Choice Questions (MCQs) through item analysis: An assessment of the assessment tool

Singh J.P.*, Kariwal P., Gupta S.B., Shrotriya V.P.

Dept. of community medicine, SRMS IMS, Bareilly (UP), India

Correspondence Address: *Dr. Jai Prakash Singh, Plot no-166, Saket nagar, Post office-Lanka, Distt-Varanasi, Pin code-221005, UP India

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 counseling or modifying learning methods.

Methods: The study was conducted during the month of November 2014 in the department of Community Medicine, SRMS Institute of Medical Sciences, Bareilly (UP). Eighty two 3rd year MBBS students took the MCQs test comprising of twenty questions. There was no negative marking and evaluation was done out of twenty marks and 50% score was the passing mark. Post validation of the paper was done by item analysis. Each item was analyzed for Difficulty index and Discrimination index.

Results: Difficulty index of 11(55%) items was in the acceptable range (p value 30-70%), 9(45%) items were too easy (p value >70%) and no any items were too difficult (p value <30%). Discrimination index of 10 (50%) items was excellent (d value>0.35), 4(20%) items was good (d value 0.20-0.34) and 6(30%) items were poor (d value<0.2%). Inter- relationship between these indices was analyzed.

Conclusion: This study inferred that items having average difficulty and high discriminating Index should be incorporated into future tests to improve the test development and review.

Keywords: Item analysis, Assessment, Multiple choice questions, Difficulty index, Discrimination index

Introduction

Assessment is an important component of a teaching-learning curriculum. A significant application of assessment 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.

Multiple-choice Questions (MCQ) are widely used for MBBS students in colleges as classroom tests and as entrance test for under-graduate and post-graduate courses. A typical MCQ item consists of a question (stem) and a set of options that consist of possible answers to the question with single best correct answer. A student's task is to select the one option that provides the best answer to the question asked. A distinct advantage of using MCQ items on classroom tests is that grading tends to be quick and without subjective bias of evaluator. Another important advantage is that a well-constructed MCQ test can yield test scores at least as reliable as those produced by a constructed-response test, while also allowing for broader coverage of the topics covered in a course (Bacon, 2003).[6]

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 counseling 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.

Aims and objective

1: To check the quality of MCQ items on the basis of responses of students.

2. To identify properly framed questions and questions those need modifications.

3. To prepare the question bank of properly framed MCQ items.

Materials and methods

This study was conducted in the department of Community medicine, SRMS Institute of Medical Sciences, Bareilly (UP) as a formative assessment during the month of November 2014. One small topic of Community medicine subject was declared 1month prior to the actual date of seminar. The students were given time to read & prepare. There was total 103 students of 3rd year (7th semester) MBBS students (93 fresh batch + 10 supplementary batch), but 21 students (16 fresh batch + 5 supplementary batch) were absent so the number of students were 82 for this study. These 82 students took the MCQs test comprising of twenty questions with single best response. There was no negative marking and the time allotted was 20 minute. Pre-validation of the paper was done by scrutinization by the faculty members of Department. Evaluation was done out of twenty 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 analyzed 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× 2/N

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

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.

Observation and result

No. of questions: 20
No. of students: 82

As seen in Table 1, there were no any questions with Difficulty Index (p) <30% so they not required modification, but Difficulty Index (p) >70% was 9 so they required modification before they can be considered as standard questions. 11 questions were within acceptable range of Difficulty Index (p) and out of those 11, 4 questions were of Difficulty Index (p) between 50% and 60%. So they can be considered as optimum as far as difficulty is concerned.

As seen in Table 2, Discrimination Index (DI) of 6 out of 20 questions was below 0.20 and hence unacceptable. DI of rest of the questions were >0.20 and so acceptable with 4 questions were categorized as having Good Discrimination (DI= 0.25 to 0.34). 10 questions were categorized as having Excellent Discrimination (DI=≥0.35).

Table 1: Difficulty Index (p)

Range	No. of Qs	Percentage of Qs
<30% (Too difficult & Unacceptable)	0	0
30%-70% (Acceptable)	11	55
>70% (Too easy & Unacceptable)	9	45
50%-60% (Ideal/Optimum)	4	20

Table 2: Discrimination Index (DI)

Range	No. of Qs	Percentage of Qs
<0.20 (Unacceptable)	6	30
0.20-0.24 (Acceptable)	0	0
0.25-0.34 (Good discrimination)	4	20
≥0.35 (Excellent discrimination)	10	50

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 The study conducted by Mehta G showed that p value of 31 (62%) items was in the acceptable range (30-70%), 16(32%) items >70% and 3(6%) items <30%.[9]

Our findings similar with the Mehta G study having p value of 11 (55%) items was in the acceptable range (30-70%), 9 (45%) items >70% and 0 (0%) 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.[11]

In the earlier study done by Mehta G, 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%), while in the present study, the items with DI > 0.35 were 10(50%), DI between 0.2 and 0.34 were 4(20%) and DI <0.2 were 6(30%). So that finding of this study was closer to study of Mehta G.

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.[12] 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.

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 4(20%) items as ideal as compared to 12(24%).[9]

Conclusion

It can be concluded from the present research that Difficulty Index (p) and Discrimination Index (DI) are very nice tool for the assessment of the quality of an MCQ item and to assess class performance as a part of formative assessment. An MCQ item should be considered unacceptable and modified to get difficulty level and discrimination power within acceptable range before it can be included in standard MCQ bank. Item analysis helps tremendously to achieve better teaching, better learning and in the long term better tests.[13]

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. Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and short answer question in marketing context. Journal of Marketing Education.25, 31-36.
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. 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.
10. Carroll RG. Evaluation of Vignette-type examination items for testing medical physiology. Am J Physiol 1993;264: S11-5.
11. 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.
12. Patel KA, Mahajan NR. Itemized Analysis of Questions of Multiple Choice Question (MCQ) Exam. International Journal of Scientific Research 2013;2:279-280.
13. 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.