Validation of Multiple Choice Questions used for evaluation of undergraduate medical students using item analysis

Pattnaik S, Selvaraj K, Mohan A

ABSTRACT

Background: Assessment quality of today's medical students is an important determinant in ensuring quality healthcare provision for tomorrow. Multiple choice questions (MCQ), is one such assessment tool, increasing in popularity in the medical field.

Aim: To validate the MCQs administered in the pre final year MBBS students during their model exam by "Item analysis."

Methods: Study was administered on 135 pre-final MBBS students as part of their model examination, which as per University guidelines contains 40 one best response type MCQs in two papers. The students were divided into two groups, viz., high achievers (top 27%/37 students) and low achievers (bottom 27%/37 students) based on their score. Item analysis of MCQs was done on three parameters; namely difficulty index (p-value), discrimination index (DI) & distracter effectiveness (DE).

Results: The mean P-value was 43 ± 17. Majority (77.5%) of the items were within recommended $(30-70)$ difficulty levels. The mean DI was 0.25 ±0.15. Most (67.5%) of the items had a 'good DI', which is a DI of more than 0.2. Only 12 items had a non-functional distracter (NF-D). Out of this 11, had one NF-D and only one had two NFD.

Conclusion: This study concludes that most of the MCQs in the exams were of average difficulty and high discrimination with functional distracters

Key Words: MCQs, validation, item analysis

INTRODUCTION

The road to reach the ultimate goal of providing quality healthcare services starts from the production of quality healthcare providers, which could be influenced greatly by the way they are assessed and this assessment in-turn depends on the quality of the assessment tool used. Due to the increasing importance being given to objectivising evaluation, multiple choice questions (MCQ) is increasingly being recognized as one of the best tools for assessing students who in turn are the future healthcare providers. There are many forms of MCQs out of which the one best response (OBR) type is a simple and efficient tool for assessing the students as well as providing feedback to teachers on their educational methods; but, merely using MCQs will not make the test more valid although it will make the test more objective.[1] Having constructed and assessed a test using MCQs, a teacher needs to know how good the test questions are.[2] Item analysis is a simple and valuable procedure performed, after the test, that provides information regarding the reliability and validity of a test item (question).[3] Item analysis usually includes three important parameters of a MCQ after it has been administered to a group of students. These parameters are difficulty index, discrimination index and distracter effectiveness.[4] The study was done with an objective of using item analysis to validate the MCQs, which were used to test the pre final year MBBS students during their model exam.

MATERIALS AND METHODS

The pre-final MBBS students were selected for the study. These students will appear in the university examination (3rd Professional Part-1) in the month of February 2014. Before they

appear in the university examination the department conducts a model exam for them in the pattern of university examination. This model examination includes two papers; each paper has 20 MCQs each along with long essay type questions and short notes. 135 students were tested each with 40 questions. All these questions were of one best response type and the time duration was 20 minutes for 20 questions. These were according to the guidelines laid down by the university.

Item analysis

For calculation of indices, the students will be divided into two groups, viz., high achievers and low achievers. After arranging the students in the descending order of their final marks in MCQs the top 27% and the bottom 27% of the students were taken as high and low achievers respectively.Three indicators were used for validation of a MCQ. They are difficulty index, discrimination index and distractor effectiveness.[4]

1. The difficulty index, item difficulty or facility value indicated by the symbol 'P' is calculated by the formula.

$$P = [(H + L)/N] \times 100$$

Where H = No. of students answering the item correctly in the high achieving group

L = No. of students answering the item correctly in the low achieving group

N = Total no. of students in the two groups (including non responders)

An item was considered difficult when the difficulty index was less than 30% and considered easy when the index was more than 70% and the value between 30-70% was acceptable (between 50-60% is ideal).

2. The discrimination index (DI) was calculated using the formula

$$DI = [(H-L)/N] \times 2$$

where the symbols H, L and N represent the same values as mentioned before.

Items with a discrimination index between 0.25 to 0.35 are good, those with more than 0.35 are excellent and those with an index below 0.2 are poor.

3. The distractor effectiveness (DE) or functionality.

Any distractor that has been selected by less than 5% of the students was considered to be a non-functional distracter. (NF-D). Distracter effectiveness for each item is calculated as follows. Four NF-D: DE = 0%; 3 NF-D: DE = 25%; 2 NF-D: DE = 50%; 1 NF-D: DE = 75%; No NFD:

DE = 100%. All data were expressed as mean and standard deviation.

RESULTS

The students were ranked based on their total scores. The top 27% (37 students) were taken as high achievers and the bottom 27 %(37 students) were taken as low achievers for analysis. The three indicators for item analysis, namely P-value, DI and DE were analysed for each MCQ.

The mean P-value for all students was found to be 43 ± 17. Majority (77.5%) of items were of average (recommended) difficulty level (Table No.1). The mean discrimination index (DI) was 0.25 ±0.15. Most (67.5%) of the MCQs were having a DI of more than 0.2 (Table No.1).One

MCQ had a minus DI (No.29), with a very difficulty level with P-value being 5. Only 12 items had a NF-D. Out of this 11, had one NF-D and only one had two NF-D (Table No.1).

Table.1. Classification of items according to the indicators

P value	Interpretation	No. of items (n=40)
>70	Too easy	03 (7.5%)
30-70	Average	31 (77.5%)
< 30	Too difficult	06 (15%)
DI	Interpretation	No. of items (n=40)
< 0.2	Poor	13(32.5%)
>0.2	Good	26 (67.5%)
		One minus
DE	Interpretation	No. of items (n=40)
100%	No NFD(Non-functional Distracter)	28 (70%)
75%	One NFD	11 (27.5%)
50%	Two NFD	01 (2.5%)
25%	Three NFD	00 (0%)
0%	All Four NFD	00 (0%)

DI: Discrimination Index ; DE: Distractor effectiveness; NFD: Non functional distractor

DISCUSSION

Any level of cognition can be assessed using a one best response MCQ, provided they are well constructed.[5] A defectively prepared MCQs will do more harm than good.[6] The difficulty index (P) and discrimination index (DI) along with the distracter effectiveness can be used as a tool in the assessment of MCQs after they have appeared in the test.

Any difficulty index between 30% -70% is acceptable with DI less than 30% indicate very difficult MCQ and a score more than 70% indicates very easy MCQ.[4] In the present study, 77.5% MCQs were found to be of acceptable difficulty level. This result is consistent with the findings of another study in Pakistan by

Mozaffer and Farhaan who found a majority of their MCQs (78%) in this acceptable range.[7] Only 7.5% and 15% MCQs in our study were too difficult and too easy respectively.

An MCQ with discrimination index above 0.2 is expected to be ably distinguished between good and weak students.[5] Our results showed that, 67.5% of the MCQs had a DI of more than 0.2. This indicates that most questions are satisfactory as they are able to differentiate good and weak students. The findings are almost similar to other study where 76% of questions had a DI of more than 0.2.[7]

Out of 160 distracters, only 13(8%) were found to be non-functional. Twelve (30%) MCQs had NF-Ds, while 28(70%) had effective distracters. This is higher than that reported (58% NF-Ds) in the study done in Pakistan.[7]

This study concludes that most of the MCQs in the exams were of average difficulty and high discrimination with functional distracters.

AUTHOR NOTE

Satyajit Pattnaik, Assistant Professor, Community Medicine ; (**Corresponding Author**);
Email: drsatyajitpattnaik@gmail.com
Kokila Selvaraj, Professor ,Community Medicine,
Arun Mohan, Assistant Professor, Community Medicine
Meenakshi Medical College and Research Institute, Kancheepuram (TN)

REFRENCES

1. Tan LT, McAleer JJ; Final FRCR Examination Board. The introduction of single best answer questions as a test of knowledge in the final examination for fellowship of the Royal College of Radiologists in Clinical Oncology. *ClinOncol (R CollRadiol)*. 2008; 20: 571-6.

2. Bharti N Karelia, AjithaPillai, Bhavisha N Vegada. 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 M.B.B.S students. *IeJSME* 2013;7(2):41-46.

3. 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.

4. Ananthakrishnan N. Item Analysis validation and banking of MCQs. In:Ananthakrishnan N, Sethuraman KR, Kumar S, editors. Medical Education Principles and Practice. 2nd ed. JIPMER, Pondicherry: p.132-33.

5. Carneson J, Delpierre G, Masters K. Designing and managing MCQs: Vol. 62, No. 2, February 2012 146.Appendix C: MCQs and Bloom's taxonomy. (Online) 2011 (Cited 2013 Dec 2). Available from http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html.

6. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. *Med Educ.* 2008; 42: 198-206.

7. Mozaffer Rahim Hingorjo and FarhanJaleel. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. *J Pak Med Assoc*. 2012;62(2): 142-47.