Item Analysis to Identify Quality Multiple Choice Questions

ARDRA RAVINDRANATHAN MENON, PRITHI NAIR KANNAMBRA

ABSTRACT

Introduction: Multiple Choice Questions (MCQs) or items are commonly used to provide feedback to the teachers at the end of an academic session. However, item analysis has to be done to ensure their quality. Since, no item analysis has ever been conducted in our department, the study was done to evaluate the quality of MCQs and create a viable question bank.

Aim: To evaluate the quality of MCQs used so as to develop a pool of valid items to update the question bank.

Materials and Methods: Total 20 items and 80 distracters from an internal examination in Microbiology, of 120 students was analysed by assessing the Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE).

Results: Out of the 20 items, 15 had acceptable DIF I (30 - 60%)

and 17 had "acceptable to excellent" DI (> 0.20). Mean DE was 87.5% considered as ideal/acceptable and Non-Functional Distracters (NFD) were only 12.5 %. Mean DI was 0.37. Poor DI in 3 items (< 0.2) with negative DI in one item indicated under preparedness of students or poor framing of at least some of the MCQs. An item becomes easier, when the number of NFDs (incorrect alternatives selected by < 5% students) in an item increases and proportionately the DE decreases. There were 8 items with 10 NFDs, while rest of the items did not have any NFD.

Conclusion: Study emphasizes the selection of quality MCQs of average difficulty and high discriminating power with functional distracters to differentiate the students in correct manner.

INTRODUCTION

Objective evaluation has become a very important tool in today's education system. MCQs or "items" are frequently used to assess students in various examinations all over the world for their objectivity and wide reach of coverage in less time. MCQs are mainly used as an excellent comprehensive tool to provide feedback to the teachers at the end of an academic session. A good and reliable MCQ can assess higher cognitive functions like interpretation, synthesis and application of knowledge [1]. Type A MCQ or the five choice type is the most commonly used one [2]. Item analysis is the process of collecting, summarizing and using information from students responses to assess the quality of test items [3]. Based on DIF I or p-value, DI and DE, the item analysis helps us to identify quality MCQ [4,5]. As MCQs are widely used for academic assessment, the present study has been done with an objective to evaluate valid MCQs by assessing with DIF I, DI, and DE. Our aim was also to create a viable question bank after revising/storing or discarding the items based on obtained results. Such an evaluation can also identify the low performers and their learning problems which can be corrected by proper counselling or by modifying our teaching 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 present cross-sectional study was conducted in the Department of Microbiology, Government Medical College, Thrissur, India, between September to November 2014, as an internal assessment after 20 hours teaching of Virology topics. Total 120 out of the total 130 second year MBBS students took the MCQs test comprising of twenty questions of type 'A' with single best response. The time for answering the questionnaire was thirty minutes. Each correct response was awarded one mark and there was no negative marking. The maximum possible overall score was twenty and 50% was considered as the pass mark. Pre-validation of the paper was done by the Head of Department and post-validation by item analysis.

Data Analysis

Data was entered and analyzed in MS excel 2007.

All the 130 second year MBBS students in the Microbiology Department were included in the study. As 10 students were absent for the evaluation, the final sample size considered

to identify quality MCQ was 120. As there was no negative marking, every student attempted all the questions. Out of the 120 students, 40 students with high scores and 40 with lower scores were taken as Higher (H) and Lower group (L) respectively, after arranging the scores in descending order. The middle 40 students were excluded from the analysis with the assumption that they behave in the similar pattern [4,6,7].Total 20 MCQs and 80 distracters were analyzed and various indices like DIF I, DI, DE, and NFD were calculated with following formulas [3-6,8]

1. DIF I or p-value = $[(H + L)/N] \times 100$ and

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

Here, N = total number of students in both high and low groups (including non-responders) and H and L are the number of correct responses in high and low groups, respectively.

Difficulty Index (DIF I or p) describe the percentage of students who selected the correct response and ranges between 0 and 100%. Bigger the value of DIF I, easier is the item and vice-versa. In general, Items with DIF I less than 30% are considered as difficult, between 30-70% are considered as acceptable, and greater than 70% are considered as easy.

Discrimination Index (item effectiveness-DI) indicates how well the question separates the students who know the material well from those who don't. The DI ranges from -1 to +1. DI of 1 is considered as ideal, which can efficiently discriminate between high and low achievers [3,4,8]. An item having a DI greater than 0.35 is considered to have excellent discriminative power and between 0.2 and 0.35 with acceptable discriminative power. An item having discrimination index '0' cannot discriminate between two (H&L) groups. An item having negative discrimination index ranging from -1 to 0 has poor discriminative power.

An ideal item has a stem and five options which includes one correct and four incorrect (distracter) alternatives. Distracter Efficiency (DE) analysis, whether the options (alternatives) were effective or not. NFD is any distracter, which has been selected by less than 5% of the students. [3,4,6,9]. DE ranges from 0 - 100% and is determined on the basis of the number of NFDs in an item. If an item contains-four or three or two or one or nil NFDs then DE will be 0, 25, 50, 75 and 100%, respectively [4,6,9,10].

RESULTS

Total 20 MCQs and 80 distracters were analyzed. Mean and Standard Deviations (SD) for DIF I (%), and DI were 44.8± 17.13% and 0 .37 ± 0.18 respectively. Out of the 20 items, 15 (75%) had acceptable level of difficulty (DIF I=30-70%), of which 10 (DIF I=50-60%) had good level of difficulty [TableFig-1].

Seventeen (85%) out of the total 20 test items had DI level of 0.2 or higher and were able to differentiate between good and weak students [Table/Fig-2].

Cut-off Value DIF I*	Items N= 20	Interpretation	Action
30- 70%	15	Acceptable	Store
<30%	5	Difficult	Revise/discard
>70%	Nil	Easy	-

[Table/Fig-1]: Distribution of items in relation to DIF I and actions proposed.
*DIF I- Difficulty Index

Cut-off Value DI*	Items N= 20	Interpretation	Action
<0 .20	3	Poor	Revise/discard
0.20 to 0.34	5	Acceptable	Store
≥0.35	12	Excellent	Store

[Table/Fig-2]: Distribution of items in relation to DI and actions proposed.
DI- Discrimination Index*

When these two were considered together, there were 13 items as ideal which could be included in question bank. However, if only the items with acceptable DIF I and excellent DI (≥0.35) were considered, there were 11 items as ideal. Besides this, one item had negative DI.

An item becomes easier, when the number of NFDs (incorrect alternatives selected by < 5% students) in an item increases and proportionately the DE decreases. There were 8 items with 10 NFDs, while rest of the items did not have any NFD [Table/Fig-3].

Distracter analysis (DE)	
Number of Items	20
Total Distracters	80
Functional Distracters	70 (87.5%)
Non functional Distracters (NFDs)	10 (12.5%)
Items with 2 NFDs (DE=50%)	2
Items with 1 NFD (DE=75%)	6
Items with 0 NFD (DE=100%)	12
Overall mean DE	87.5±17.2%

[Table/Fig-3]: Distracter analysis (N=100).

DISCUSSION

The effective measurement of knowledge, skill and competence acquired is an important component of medical education. Though, MCQs are constructed to test simple, 'factual recall', careful construction can also evaluate higher order of thinking skills like comprehension, application, analysis, synthesis and judgement, which is very important for a medical graduate. Post examination item analysis with DIF I, DI and DE is a simple but effective method to assess the validity and reliability of a test, detects specific technical flaws and provides information for further improvement [11].

Out of the 20 MCQs given to 120 students, the mean DIF I

Mean ±SD	Mehta G et al., 2014 [1]	Gajjar S et al., 2014 [4]	Hingorjo MR et al., 2012 [6]	Kaur M et al., 2016 [7]	Pande SS et al., 2013 [8]	Present study
Difficulty Index I	63.06±18.95	39.4±21.4	54.14±17.48	48.90±13.72	52.53±20.59	44.8±17.13
Discrimination Index	0.33±0.18	0.14±0.19	0.356 ± 0.17	0.37±0.15	0.30±0.18	0.37±0.18
Distractor Efficiency %	63.97±33.56	88.6±18.6	81.41	83.98±24.52	Not done	87.5±17.2

[Table/Fig-4]: Comparing data of similar studies of item analysis with the present study.

scores of the individual tests in this study was 44.8± 17.13, well within the acceptable range (30- 70%). Previous studies have proposed the mean of DIF I as 39.4± 21.4, 48.90±13.72 and 52.53±20.59 [4,7,8] [Table/Fig-4]. Since, DI refers to the percentage getting the item right, the smaller percentage figures the more difficult item. So items with DIF 1< 30 % will be more difficult and lead to low scores while those with DIF > 70% will be easier and give a false blanket of confidence. In the present analysis, 5 MCQs (25%) with DIF I <30% should be reviewed for any confusing language, controversies or even an incorrect key and revised or else discarded. When the DFI I is very small as in our 1st question with DIF I=6.25, indicating difficult question, it may be that the test item is not taught well or is difficult for the students to grasp. It also may indicate that the topic tested is inappropriate at that level for the students. Difficult questions were discussed with students which helped them to clear their doubts. In this study, there was no easy question (DIF I >70%). It is advisable to place easy questions also in test paper to increase the confidence of different types of students. Similarly, difficult questions can be retained and used to select toppers.

DI of an item discriminates between higher and lower achievers. Mean DI in present study was 0.37±0.18 which is considered excellent according to the cut-off point of 0.2 which is comparative to other studies [Table/Fig-4].Three items (15%) showed DI <0.2, with negative DI in one case. It is obvious that a question which is either too difficult or too easy will have nil to poor DI. Whether to retain such item depends on its relevance. Mean DI in a study by Gajjar S et al., was 0.14±0.19 less than the acceptable cut off point of 0.2, because 10 out of the 50 items had negative DI [4]. Negative DI can be due to incorrect key, poor framing of question or generalized under preparation of students. In the present scenario, it was a question from nice to know portion of syllabus and the students were under prepared for it. The validity of a test can be decreased by items with negative DI.

DIF I and DI are usually reciprocally related but their relationship is often considered dome shaped and not linear. Questions having high DIF I value (easier questions) discriminate poorly and vice-versa except where DIF I is either too high or low [12]. Based on the cut off points for "acceptable to excellent" for DIF I and DI (DIF I value = 30 to 70; DI > 0.2), there were 13 items (65%) considered ideal. Out of the 13 tests, 11(84.6%) showed DI equal to or more than 0.35. These MCQs can be considered excellent test items to differentiate students of higher and lower abilities.

Writing appropriate options to the correct answer is the most difficult task in the construction of a good MCQ. Relative use of distracters in each item can be assessed by distracter analysis. The mean DE was 87.5±17.2. There were 8 items with 10 NFDs (9%). In a study by Gajjar et al., among 150 distracters, 133 (89.6%) were functional and 17 (11.4%) were NFDs. [4]. In another study, with 120 distracters, 91 (75.8%) were functional distracters, and 29 (24.16%) were NFDs [10]. NFDs should either be removed or replaced with a more plausible option. More NFDs in an item increases DIF I (makes item easy) and reduces DI. In the present study, 6 out of the 8 items with NFDs had DIF 1 >50%. So, in the framing of a quality MCQ, writing plausible distracters and reducing the NFDs is very important. However, NFD is not the only factor that contributes to the difficulty of an MCQ item. Flaws in item writing also contribute to poor student performance. MCQs in medical education are usually constructed by doctors who have other commitments like clinical work and administrative responsibilities. So training faculty in the item writing skills along with regular pre and post-validation of MCQ items is essential [13].

LIMITATIONS
Our test contained only 20 questions, the tentative nature of our data requires that we follow for a wide margin of error. Since, this is the very first time that we have preceded with item analysis of our MCQs, we need a continuous evaluation at the end of all sessions to create a viable question bank.

CONCLUSION
An item with average difficulty (DIF I 30-70%), high discrimination (DI ≥ 0.2) and maximum DE (100%) is considered as an ideal MCQ. Discussion about results of item analysis with faculty as well as with students helps in improving learning outcome. If an item provides a positive index for discrimination, if all the alternatives are functioning effectively, and if the item measures an educationally significant outcome, it should be retained in an item file for future use. Periodic review of the items in the question bank after each examination will identify the areas of potential weakness, make the information more interpretable and create an ideal item bank.

ACKNOWLEDGEMENTS

REFERENCES

[1] 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.

[2] Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. IeJSME. 2009;3(1):02-07.

[3] Singh T, Gupta P, Singh D. Test and Item analysis.In: Principles of Medical Education.4th ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd.2013; p 109.

[4] 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 Ahmedabad, Gujarat. Indian J Community Med. 2014;39:17-20.

[5] Hand book of Manual of Basic Workshop in medical education technologies: Thrissur Govt. Medical College; 2014.

[6] Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc. 2012;62(2):142-47.

[7] Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. Int J Appl Basic Med Res. 2016;6(3):170-73.

[8] Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty & discrimination indices of MCQs in formative exam in physiology. South East Asian J Med Educ. 2013;7:45–50.

[9] Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Med Educ. 2009;9:40.

[10] Kolte V. Item analysis of multiple choice questions in physiology examination. Indian J Basic Appl Med Res. 2015;4:320–26.

[11] 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.

[12] 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(2):67-71.

[13] Abdulghani HM, Ahmad F, Ponnamperuma GG, Khalil MS, Aldrees A. The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. J Health Spec. 2014;2:148-51.

AUTHOR(S):
1. Dr. Ardra Ravindranathan Menon
2. Dr. Prithi Nair Kannambra

PARTICULARS OF CONTRIBUTORS:
1. Assistant Professor, Department of Microbiology, Government Medical College, Thrissur, Kerala, India.
2. Professor and Head, Department of Microbiology, Government Medical College, Thrissur, Kerala, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Dr. Ardra Ravindranathan Menon,
Assistant Professor, Department of Microbiology,
Government Medical College,
Thrissur-680596, Kerala, India.
E-mail: drardra@yahoo.co.in

FINANCIAL OR OTHER COMPETING INTERESTS: None.

Date of Publishing: **Apr 01, 2017**