International Journal of Research in Medical Sciences
Sahoo DP et al. Int J Res Med Sci. 2017 Dec;5(12):5351-5355
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20175453

Original Research Article

Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students

Durgesh Prasad Sahoo, Rakesh Singh*

Department of Community Medicine, IGGMC, Nagpur, Maharashtra, India

Received: 26 September 2017
Accepted: 31 October 2017

***Correspondence:**
Dr. Rakesh Singh,
E-mail: docrakesh06@gmail.com

ABSTRACT

Background: Multiple choice questions (MCQs) or Items forms an important part to assess students in different educational streams. It is an objective mode of assessment which requires both the validity and reliability depending on the characteristics of its items i.e. difficulty index, discrimination index and distracter efficiency. To evaluate MCQs or items and build a bank of high-quality test items by assessing with difficulty index, discrimination index and distracter efficiency and also to revise/store or remove errant items based on obtained results.
Methods: A preliminary examination of Third MBBS Part-1 was conducted by Department of Community Medicine undertaken for 100 students. Two separate papers with total 30 MCQs or items and 90 distractors each in both papers were analyzed and compared. Descriptive as well as inferential statistics were used to analyze the data.
Results: The findings show that most of the items were falling in acceptable range of difficulty level however some items were rejected due to poor discrimination index. Overall paper I was found to be more difficult and more discriminatory, but its distractor efficiency was slightly low as compared to paper II.
Conclusions: The analysis helped us in selection of quality MCQs having high discrimination and average difficulty with three functional distractors. This should be incorporated into future evaluations to improve the test score and properly discriminate among the students.

Keywords: Difficulty index, Discrimination index, Distractor efficiency, Functional distractor, Non-functional distractor

INTRODUCTION

Multiple choice question (MCQ) assessments are becoming more and more prevalent for assessing knowledge for many professional courses including Medicine. They provide faster ways of assessing student learning and can be used effectively to measure a wide range of abilities. It is a time-tested method of assessment of knowledge in both undergraduate and postgraduate medical education for the purpose of ranking in the order of merit.[1,2]

Item analysis is a process, where we use to examine response of the student to individual test items or questions to assess the quality of those items and of the test as a whole. It not only helps in improving items which will be used again in later tests, but also it can also be used to eliminate ambiguous or misleading items in a single test administration.

In addition, item analysis is also valuable for increasing instructors' skills in test construction, and identifying specific areas of course content which need greater emphasis or clarity.[3]

Properly constructed multiple choice questions assess higher-order cognitive processing such as interpretation, synthesis and application of knowledge, instead of just

testing recall of isolated facts and is preferred over other methods for its (a) objectivity in assessment, (b) minimization of assessor's bias, (c) precise interpretation for content validity, (d) assessing a diversity of content, and (e) can be used with all subject areas. Item analysis enables identifying good MCQs based on difficulty index (DIF I) also denoted by FV (facility index), discrimination index (DI), and distractor efficiency (DE).[4,5]

Item analysis is an efficient tool in identifying the strengths and weaknesses in students, as well as providing guidelines to teachers for preparing good MCQs. Multiple Choice Question (MCQ) based assessments are very common nowadays, yet not enough stress is laid on the training of teachers on preparing good MCQs. There are very few studies from India reporting use of item analysis of MCQs in Medical education.

Hence, we present this simple analysis tool used for MCQs which can help improve the quality of MCQ-based assessments with the aim to find out high-quality test items based on the relationship of items having good difficulty and discriminator indices, with high distractor efficiency. This was achieved through evaluating MCQs or items and build a bank of high-quality test items by assessing with DIF I, DI and DE and also to revise/store or remove errant items based on obtained results.

METHODS

This study was performed on two preliminary examination papers (Paper I and Paper II) containing 30 MCQs each, faced by 100 Third MBBS Part-1 undergraduate students in Community Medicine, from a government medical college, Mumbai in 2014. The paper included single best response type MCQs, with four choices. Each correct response was awarded half mark. No mark was given for blank response or incorrect answer. There was no negative marking. Thus, the maximum possible score of the overall test was 15 and the minimum 0.

Data obtained was entered in MS Excel 2013 and analysed score of 100 students was categorized into the high scoring (H) group (top 33%), mid scoring (M) group (middle 34%) and the low scoring (L) group (bottom 33%) respectively, after arranging the scores in descending order.[6]

So out of 100 students, 33 were in H group and 33 in L group; rests (34) were in middle group and not considered in the study. Total 30 MCQs and 90 distractors for each paper were analysed and based on this data, various indices like DIF I, DI, DE, and non-functional distractor (NFD) were calculated for each item as follows:

DIF I = [(H+L)/N] x 100
DI = [(H-L)/N] x 2

Where, N is the total number of students in both the groups. H and L are number of correct responses in the high and low scoring groups, respectively.[7]

Difficulty index also called ease index, describes the percentage of students who correctly answered the item. It ranges from 0-100%. The higher the percentage, the easier the item and vice versa. The recommended range of difficulty is from 30-70%. Items having DIF I below 30% and above 70% are considered difficult and easy items respectively.

Discrimination index (DI) explains the ability of an item to differentiate between high and low scorers. It ranges between -1.00 and +1.00. Item with higher value of DI is better able to discriminate between students of higher and lower abilities.

Discrimination index is classified as

- 0.40 And above: very good items,
- 0.30–0.39: reasonably good,
- 0.20–0.29: marginal items (i.e. subject to improvement), and
- 0.19 Or less is poor items (i.e. to be rejected or improved by revision).[8]

Ideal DI is 1, as it refers to an item which perfectly discriminates between students of lower and higher abilities.[9] The high-performing students can select the correct answer for each item more often than the low-performing students. If this is true, then the assessment is having a positive DI (between 0.00 and +1.00). This indicates those students who received a high total score will choose the correct answer for a specific item more times than the students who had a low overall score. However, if the low performing students will get a specific item correct more times than the high scorers, then that item will have a negative DI (between -1.00 and 0.00). Here a good student suspicious of an easy question, takes harder path to solve and end up being less successful while a student of lower ability by guess select correct response.[10]

The difficulty index (DIF I) and discrimination index (DI) are often reciprocally related. However, this may not be true always. Questions having high DIF I (easier questions), discriminate poorly as compared to a question with a low DIF I (harder questions), which are considered to be good discriminators.[11]

An item i.e. MCQ contains a question which presents a problem situation and four options i.e. one correct (key) and three incorrect (distractor) alternatives.[10] Non-functional distractor (NFD) in an item is option (other than key) selected by <5% of students; alternatively, functional or effective distractors are those selected by 5% or more participants.[12,13] Distractor efficiency (DE) ranged from 0-100% and was determined on the basis of the number of NFDs in an item. 3 NFD: DE = 0%; 2

NFD: DE = 33.3%; 1 NFD: DE = 66.6%; 0 NFD: DE = 100%.[12,13] By analyzing the distractors, it becomes easier to identify their errors, so that they may be revised, replaced, or removed.[7]

RESULTS

A total 100 students gave the test consisting 30 MCQs and 90 distractors in each Paper I and Paper II.

Table 1: Difficulty index, discrimination index and distractor efficiency.

Parameter	Paper I		Paper II	
	Mean	SD	Mean	SD
Difficulty index (DIF I)	51.87	19.69	58.03	21.31
Discriminatory index (DI)	0.29	0.14	0.26	0.15
Distractor efficiency (DE)	84.42	19.07	86.64	18.80

Table 1 shows that paper I is more difficult and more discriminatory, but the distractor efficiency was less as compared to paper II.

Overall, both papers were having good or excellent difficulty level and distractor efficiency, but the discriminatory power was marginal.

Table 2: Classification of MCQs according to difficulty and discrimination indices and actions proposed.

Cut off points	Paper I	Paper II	Interpretation	Action
	Items (n=30)	Items (n=30)		
Difficulty index (DIF I)				
< 30	3	3	Difficult	Revise or discard
30-70	22	18	Good or excellent	Store
> 70	5	9	Easy	Revise or discard
Discriminatory index (DI)				
≤ 0.19	9	12	Poor	Revise or discard
0.20-0.29	6	6	Marginal	Revise or discard
0.30-0.39	8	7	Good	Store
≥ 0.40	7	5	Excellent	Store

Table 2 shows that: Paper I: Out of 30 items, 22 had "good to excellent" DIF I (30-70%) and 15 had "good to excellent" DI (≥0.30). Mean DIF I was 51.87 % and mean DI was 0.29.

Paper II: Out of 30 items, 18 had "good to excellent" DIF I (30-70%) and 12 had "good to excellent" DI (≥0.30). Mean DIF I was 58.03 % and mean DI was 0.26.

On comparing, Paper I have more no. of items which were good to excellent with better difficulty index and were more discriminatory as compared to Paper II.

Table 3: Distractor analysis.

Number of items	Paper I	Paper II
	30	30
Total distractors	90	90
Functional distractors	73 (81.11%)	71 (78.89%)
Non-functional distractors (NFDS)	14 (15.56%)	12 (13.33%)
No response	3 (3.33%)	7 (7.78%)
Items with 1 or 2 NFDS (DE between 66.6 and 33.3%)	13 (43.33%)	11 (36.67%)
Items with 0 NFDS (DE= 100%)	17	19
Items with 1 NFDS (DE= 66.6%)	12	10
Items with 2 NFDS (DE= 33.3%)	1	1
Overall DE (Mean±SD)	84.42±19.07	86.64±18.80

Table 3 shows that, Paper I: Mean DE was 84.42 % considered as ideal or acceptable and non-functional distractors (NFDs) were only 15.56%. Increased proportion of NFDs (incorrect alternatives selected by <5% students) in an item decrease DE and makes it easier. There were 13 items with 14 NFDs, while rest items have 0 NFDs with mean DE of 100%.[17]

Paper II: Mean DE was 86.64 % considered as ideal or acceptable and non-functional distractors (NFD) were only 13.33 %. There were 11 items with 12 NFDs, while rest items have 0 NFDs with mean DE of 100%.[19]

On comparing, distractor efficiency of Paper II was more as compared to Paper I because no. of NFDs were more in Paper I.

DISCUSSION

The assessment tool is a strategy which should be designed as per the objectives. MCQs, if properly written and well-constructed, is one of the strategy of the assessment tool that quickly assess any level of cognition as per Bloom's taxonomy.[14]

The difficulty and discrimination indices are the tools used to check whether the MCQs are well constructed or not. Another tool we used for further analysis is the distractor efficiency. It analyses the quality of distractors and it is closely associated with difficulty and discrimination indices.

An ideal item (MCQ) will be the one which has average difficulty index (DIF I between 30 and 70%), high discrimination index (DI≥0.30) and maximum DE

(100%) with three functional distractors. Assessment of MCQs by these indices explains the importance of assessment tools for the benefit of both the student as well as for the teacher.[7]

Mean DIF I for Paper I was 51.87±19.69% and for Paper II was 58.03±21.31% well within the acceptable range (30-70%) identified in present study which is comparable to similar study done by Kumar P et al among medical students in Gujarat with mean DIF I as 39.4±21.4%.[15] Another study done by Pande SS et al have proposed the mean of DIF I as 52.53±20.59%.[16] Guilbert J. proposed the range for DIF I as 41-60%.[17] Too difficult items (DIF I <30%) will lead to low scores, while the easy items (DIF I >70%) will result into high scores.

In the present study, mean DI for Paper I was 0.29±0.14 and for Paper II was 0.26±0.15 which shows that items are marginal and needs revision. This was due to 50% items have DI values <0.30 in paper 1 and 60% items have DI values <0.30 in paper 2. Similar study by Kumar P et al observed, mean DI as 0.14 ± 0.19 which was less than the acceptable cut off point of 0.30.[15] Items with DI value less than 0.20 needs to be revised or discarded. When good to excellent DIF I and DI are considered together, there were 15 items as ideal which could be included in question bank from Paper I and 12 items from Paper II (out of 30) as compared to 15 (out of 50) by Kumar P et al and 32 (out of 50) in another study by Hingorjo MR et al.[15,13]

Designing of functional distractors and also reducing the NFDs are important aspect for framing quality MCQs.[18] More NFD in an item increases DIF I and reduces DE while item with more functioning distractors decreases DIF I and increases DE. Higher the DE more difficult the question and vice versa, which ultimately relies on presence or absence of NFDs in an item. Mean DE in present study amongst Paper I was 84.42±19.07% and amongst Paper II was 86.64±18.80% which is comparable to similar type of study with DE of 88.6±18.6% and 81.4%.[13,15] Limitation of this study was, item analysis data are tentative which are influenced by the type and number of students being tested, instructional procedures employed, and chance errors.

CONCLUSION

Items analysed in the study were neither too easy nor too difficult (mean DIF I = 51.87% and 58.03%) which is acceptable. The mean DI was 0.29 and 0.26 which denotes items were marginal at differentiating higher and lower ability students and the distractor efficiency of both papers was good. This study highlights the importance of item analysis. Items having high discrimination, average difficulty level and high distraction efficiency should be incorporated into future tests so as to improve the evaluative test development and review. This would not only improve the overall test score but would also properly discriminate among the students.

ACKNOWLEDGEMENTS

Authors would like to thank all the students of third MBBS part-1 of Parent College and staff of Department of Community Medicine for their support during the study.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Hamdorf JM, Hall JC. The development of undergraduate curricula in surgery: I. General issues. Aust N Z J Surg. 2001;71(1):46-51.
2. Liyanage CAH, Ariyaratne MHJ, Dankanda DHCHP, Deen KI. Multiple choice questions as a ranking tool : a friend or foe ? J Med Educ. 2009;3(1):62-4.
3. Office of Educational Assessment. Understanding Item,2017. Available at: http://www.washington.edu /assessment/scanning-scoring/scoring/reports/item-analysis/.
4. Carneson J, Delpierre G, Masters K. Designing and Managing Multiple Choice Questions,2016. Available at: http://web.uct.ac.za/projects/cbe/mcqman/mcqman01.html.
5. Case SM, Swanson DB. The selection of upper and lower groups for the validation of test items. Director. 2002;27(21):112.
6. Kelley TL. The selection of upper and lower groups for the validation of test items. J Educational psychol. 1939;30(1):17.
7. Miller MD, Linn RL GN. Measurement and assessment in teaching. Up Saddle River, NJ Prentice Hall. 2009.
8. Garvin AD, Ebel RL. Essentials of Educational Measurement. Educ Res. 1980;9(9):232.
9. Singh T, Gupta P SD. Test and item analysis. In: Principles of Medical Education, 3rd ed. New Delhi, Jaypee Brother Med Pub Ltd. 2009:70-7.
10. Susan Matlock-Hetzel. Basic Concepts in Item and Test Analysis,1997. Available at: http://ericae.net/ft/tamu/Espy.htm.
11. Carroll G, Carolina N, Evaluation RG, Carolina E. Evaluation for testing of vignette-type examination medical physiology student. 1993;264.
12. 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(1):40.
13. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012;62(2):142-7.
14. Crooks TJ. The Impact of Classroom Evaluation Practices on Students. Am Educ Res Assoc. 2016;58(4):438-81.

15. Kumar P, Sharma R, Rana M, Gajjar S. 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(1):17.
16. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. South-East Asian J Med Educ. 2013;7(1):45-50.
17. Guilbert J, Organization WH. Educational Handbook for Health Personnel Sixth Edition. World Health. 1987.
18. Rodriguez MC. Three options are optimal for multiple-choice items: A meta- analysis of 80 years of research. Educ Meas Is Pract. 2005;24(2):3-13.

Cite this article as: Sahoo DP, Singh R. Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. Int J Res Med Sci 2017;5:5351-5.