

Research Article

Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India

Rajkumar Patil^{1*}, Sachin Bhaskar Palve¹, Kamesh Vell², Abhijit Vinod Boratne¹

¹Department of Community Medicine, Mahatma Gandhi Medical College and Research Institute, Pillaiyarkuppam, Sri Balaji Vidyapeeth University, Pondicherry 607402, India

²Department of Community Medicine, Sree Lakshmi Narayana Institute of Medical Sciences, Osudu, Agaram Village, Villianur Commune Kudupakkam Post, Pondicherry 605502, India

Received: 21 April 2016

Revised: 21 May 2016

Accepted: 21 May 2016

*Correspondence:

Dr. Rajkumar Patil,
E-mail: drraj49@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Medical students are evaluated and assessed by different methods. One of the methods for evaluation is by using multiple choice questions (MCQs). MCQs are difficult to frame but easy to administer and check. MCQs can be evaluated for the quality by item and test analysis. The objective of the study was to evaluate the MCQs among seventh semester MBBS students.

Methods: Total 30 MCQs were constructed in community medicine. These MCQs were administered to a small group of MBBS students. All MCQs had single stem with three wrong and one correct answer. The data were entered in Microsoft excel 2010 software and analysed. Mean, SD, Proportions were used. Difficulty index (DIF I), discrimination index (DI) and distractor efficiency (DE) were the parameters used to evaluate the items.

Results: Total 90 distractors (3x30 MCQs) were analysed. Mean for difficulty index, discrimination index and distractor efficiency were 38.3%, 0.27 and 82.8% respectively. Of 30 items, 11 items were of higher difficulty level (DIF I <30%) while 5 items were of easy level (DIF I >60%). Total 15 items were having very good DI. Of the 90 distractors, there were 16 (17.8%) non-functional distractors (NFDs) present in 13 (43.3%) items.

Conclusions: In present study, there were only three MCQs out of the total 30 MCQs which satisfied all the criteria for an ideal MCQ.

Keywords: MCQs, Item analysis, Difficulty index, Discrimination index, Distractor efficiency, Medical education

INTRODUCTION

Assessment is an integral part of any learning and training. Medical students are evaluated and assessed by different methods. One of the methods for evaluation is by using multiple choice questions (MCQs). MCQs are having high objectivity which avoids inter-examiner bias, these are difficult to frame but easy to administer. The results are easy to compile and analyse. Although MCQs are not commonly used in assessment of MBBS and

medical postgraduate students, these are often the choice for most of the graduate and postgraduate medical entrance examinations. MCQs can be designed to assess the higher cognitive levels of the students.

An MCQ has one item stem and possible options. Stem can be in question form or can be an incomplete statement. Mostly an MCQ (with single best answer) has four options with one correct answer and three wrong options which act as distractors.

Evaluation of MCQ is done by item analysis, it is the process of collecting, summarizing and using information of the students after conducting a test based on MCQs. It analyses the performance of an individual MCQ and the overall MCQs test.¹

Difficulty index (DIF I), discrimination index (DI) and distractor efficiency (DE) are the parameters used to evaluate the items.²⁻⁴

Evaluation of constructed MCQ is necessary for following reasons:

- to know the difficulty level (appropriate/inappropriate) of question,
- to know whether the question is able to discriminate between high and low achievers
- to know the plausibility of options other than correct answer (distractors)

Overall, item analysis provides feedback to teachers for necessary modifications in MCQs to make it suitable for the exam. While some MCQs are edited, some are deleted based on the analysis.^{5,6}

Present study is conducted with an objective of evaluation of MCQs among MBBS students.

METHODS

Present study was planned and conducted as a small project under “Basic course on Medical Education Technologies” workshop held at Mahatma Gandhi Medical College and Research Institute, Pondicherry, India. Total 30 MCQs were constructed in general epidemiology chapter of community medicine which mainly included history of epidemiology, infectious disease epidemiology, screening of diseases, measurements in epidemiology and various types of study designs in epidemiology. These questions were vetted by two of the subject experts.

While constructing MCQs following points were kept in mind:

- Single best answer for each question
- Avoidance of the following:
 - a. Absolute options e.g. never, always, all of the above, none of the above
 - b. Ambiguous options
 - c. Repetition of part of the stem in options
 - d. Double negative stem
- Wherever word “except” was used in stem it was written in capital letters (EXCEPT)
- Options were placed in a manner to avoid any particular fixed pattern of correct answers

- Acronyms were avoided in stem and options. Wherever acronym was used, its expansion was also written.

All MCQs had single stem with four options including, one correct answer and other three incorrect answers (distractors). Verbal consent was obtained from the students. These MCQs were administered to the group of 20 MBBS students of 7th semester. For all these students general epidemiology chapter was already taught during their previous semesters. For each correct answer one mark was allotted. The maximum possible score was 30 and minimum 0. There was no negative marking for wrong answers.

The data obtained was entered in Microsoft Excel 2010 and analysed. Mean, SD, Proportions were used. For evaluation of MCQs, marks of all 20 students were ranked in descending order from highest score to lowest score. After arranging the scores in descending order, three groups were made: 30% high achievers, 40% middle achievers and 30% low achievers.

Difficulty Index (DIF I), Discrimination Index (DI) and Distractor efficiency (DE) were used for evaluation of MCQs.²⁻⁴

Difficulty index (DIF I)

It is the percentage of students who select the correct answer for an item. Higher the value of difficulty Index easier is the question, so higher value of DIF I mean easy questions. It is calculated as percentage of students who correctly answered the item. It ranges from 0-100%.

DIF I was calculated by the formula: $(H+L / N) \times 100$

Where, H - Number of the students answering the item correctly in the high group

L - Number of the students answering the item correctly in the low group

N - Total number of the students in two groups including the non-responders

The difficulty Index for an item was categorized as follows:

- ≤ 30 : difficult MCQ
- 31-40: good MCQ
- 41-60: very good MCQ
- > 60 : easy MCQ

For an item DIF I of 31-60% can be considered adequate, if it is above 60% or below 31% the MCQ may require some modification.

Discrimination index (DI)

It is the ability of an item to differentiate between the high and low achievers and it ranges from 0 to 1. If DI is higher, the item is more able to discriminate between high and low achievers.⁴

Discrimination index (DI) was calculated by the formula:
 $2 \times [(H-L) / N]$

Where, H, L and N are same as above mentioned in difficulty index.

Distractor efficiency (DE)

It shows the effectiveness of the incorrect options (distractors) given in the item. It simply shows whether distractors are functioning as distractors or not functioning. Non-functioning distractor (NFD) is an option other than correct answer which is selected by less than 5% of total students in high and low group while the distractors which are selected by 5% or more than 5% of the students are considered as functional distractors.⁷

Distractor efficiency was determined for each item on the basis of the number of NFDs in it and ranged from 0 to 100%. DE was 100%, 66.6%, 33.3% and 0% based on presence of zero, one, two or three NFDs in an item respectively. An MCQ satisfying all three criterion (DIF I, DI, DE) of “good to very good MCQ” was considered as ideal.

RESULTS

Total 30 multiple choice questions (MCQs) were constructed and evaluated among 20 students. Total 90 distractors (3x30 MCQs) were analysed. Mean score and standard deviation were 11.7 and 3.7 respectively. Total score out of 30, ranged from 2 to 17 (6.7% to 56.7% marks). For evaluation, marks of students were ranked in descending order from highest score of 17 to lowest score of 2. The first 30% students (6) were included in High group and the last 30% (6) students in low group. Eight students' data in middle group were not used for the evaluation.

In present study, mean and standard deviations for difficulty index (%), discrimination index and distractor efficiency (%) were 38.3 (22.5), 0.27 (0.28) and 82.8 (22.5) respectively.

Of 30 items, 11 items were of higher difficulty level (DIF I <30%) while 5 items were of easy level (DIF I >60%). Total 14 items were middle two levels which can be considered as good to very good items (Table 1).

Total 9 items were of poor discriminatory level in which 2 items had negative DI. Total 15 items were having very good DI (Table 2).

Table 1: Distribution of items in relation to difficulty index (DIF I).

Difficulty index % (DIF I)	Items (n-30)	Interpretation
≤30	11 (36.7)	Difficult MCQ
31-40	4 (13.3)	Good MCQ
41-60	10 (33.3)	Very good MCQ
≥60	5 (16.7)	Easy MCQ

*parenthesis indicates percentage

Table 2: Distribution of items in relation to discrimination index (DI).

Discrimination index (DI)	Items (n-30)	Interpretation
<0.15	9 (30.0)	Poor MCQ
0.15-0.24	6 (20.0)	Good MCQ
≥0.25	15 (50.0)	Very good MCQ

*parenthesis indicates percentage

Of the 90 distractors, there were 16 (17.8%) NFDs present in 13 (43.3%) items. Total 10 items were having one NFD while 2 NFDs were present in each of the 3 items (Table 3).

Table 3: Distractor analysis.

Distractor analysis	Number
Number of items	30
Total distractors	90
Functional distractors	74 (82.2)
Non-functional distractors (NFDs)	16 (17.8)
Items with any NFD	17 (56.7)
Items with any NFD	13 (43.3)
Items with 1 NFD (DE 66.7%)	10 (33.3)
Items with 2 NFDs (DE 33.3%)	3 (10)
Items with 3 NFDs (DE 0%)	0

*parenthesis indicates percentage

There were less number of items with NFDs in difficult questions (DIF I up to 40%). There were 8 items with NFDs in high (>0.25) discrimination index group. Out of 13 NFD, 84.6% of the items discriminates between high and low achievers (Table 4).

Table 4: Items with non-functional distractors and their relationship with DIF I and DI.

Difficulty index % (DIF I)	Items with NFD (n-13)	Discrimination index (DI)	Items with NFD(n-13)
≤30	1 (7.6)	<0.15	2 (15.4)
31-40	2 (15.4)	0.15-0.24	3 (23.1)
41-60	5 (38.5)	≥0.25	8 (61.5)
≥60	5 (38.5)	-	-

DISCUSSION

Along with thorough understanding and in depth scrutiny of the topic, MCQs help medical personnel develop the cognitive knowledge and add more in acquiring skills to the subject. Any assessment whether formative or summative has intense effect on learning and is an important variable in directing the learners in a meticulous way. Single correct response type MCQ is an efficient tool for evaluation. The quality of MCQ is assessed by the analysis of item and test as a whole together which is referred as item and test analysis. An ideal MCQ should have average level of difficulty ($>30-60\%$) with higher discrimination index (≥ 0.25) and 100% distractor efficiency (means all three incorrect responses should function).

In present study, as per the difficulty index criteria 15 MCQ were good to very good MCQs, as per the discrimination index 21 were good to very good MCQs while on the basis of distractor efficiency 17 MCQs were ideal while there were only 3 MCQs out of the total 30 MCQs which satisfied all the criteria for an ideal MCQ. It means that quality of MCQs required for assessment was poor which indicates that it is difficult to construct an ideal MCQ. There may be possibility that time for preparation of the given topic might be inadequate.

In the present study, the means and standard deviations for DIF I (%), DI and DE (%) were 38.34 ± 22.49 , 0.27 ± 0.28 and 82.8 ± 22.5 . In a study by Sanju Gajjar et al, for item and test analysis to identify quality multiple choice questions (MCQs) in Gujarat, means and standard deviations (SD) for DIF I (%), DI and DE (%) were $39.4 \pm 21.4\%$, 0.14 ± 0.19 , and $88.6 \pm 18.6\%$ respectively. These means are nearly same as the findings of our study.⁵

In the study by Gajjar S et al, about 50% of the items had “good to excellent” level of difficulty and 50% had “good to excellent” discrimination power ($DI \geq 0.15$).⁸ The present study also showed that 46.6% items had “good to excellent” level of difficulty and about 67% had “good to excellent” discrimination power ($DI \geq 0.15$). It means that MCQs which are evaluated in other studies also need modifications almost similar to our study.

In the present study, total 9 items were of poor discriminatory level out of which 2 items had negative DI. Negative DI for an item means that the lower achievers answered that particular item more correctly compared to high achievers. It may happen sometimes when some students of lower group guess the answers correctly. Total 15 items were having very good DI which indicates that these 15 can be used to differentiate high and low achievers.

Analysing the distractors (incorrect alternatives) is done to determine their relative usefulness in each item. In the present study, all the distractors were functioning in 17

(56.7%) MCQs, it means distractor efficiency (DE) was 100% in these questions. In a study by Tarrant M et al, the proportion of items containing all three functioning distractor was 13.8%.⁷ Items need to be modified if students consistently fail to select certain distractors. Such distractors are probably implausible and therefore of little use as decoys. Many times examiners face difficulty in developing three or more equally plausible distractors. It is better to have an item with two plausible distractors rather than an item with three or four implausible distractors.⁹

Therefore, designing of plausible distractors and reducing the NFDs is an important aspect for framing quality MCQs. More number of non-functional distractors in an item increases DIF I (makes item easy) and reduces DE, conversely item with more functioning distractors decreases DIF I (makes item difficult) 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.

Limitation of current study is that it was conducted in a small group of students with small number of MCQs but the current study was more focussed on the evaluation method of MCQs which should be used in any assessment of the students based on MCQs.

CONCLUSION

In present study, there were only 3 MCQs out of the total 30 MCQs which satisfied all the criteria for an ideal MCQ. Development of an MCQ requires more efforts keeping in mind the qualities of an ideal MCQ.

Recommendations

It is must to perform item analysis of MCQs in order to make quality MCQs. Different group of medical students may perceive the difficulty level differently, so it is better to administer the MCQs to a large number of students, to modify the questions appropriately.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

1. 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 I multidisciplinary summative tests. International e-Journal of Science, Medicine and Education. 2009;3(1):2-7.
2. Sarin YK, Khurana M, Natsu MV, Thomas AG, Singh T. Item analysis of published MCQs. Indian Pediatrics. 1998;35:1103-5.
3. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and

- distractor efficiency. *Journal of Pakistan Medical Association*. 2012;62:142-7.
4. Singh T, Gupta P, Singh D. *Principles of Medical Education*. 3rd Ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd. 2009. Test and item analysis; p.70–7.
 5. Zubairi AM, Kassim NL. Classical and Rasch analysis of dichotomously scored reading comprehension test items. *Malaysian Journal of English Language Teaching Research*. 2006; 2:1-20.
 6. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals Academy of Medicine Singapore*. 2006;35:67-71.
 7. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*. 2009;9:1-8.
 8. 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 Journal of Community Medicine*. 2014;39(1):17-20.
 9. Schuwirth LWT, Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education*. 2004;38(9):974-9.

Cite this article as: Patil R, Palve SB, Vell K, Boratne AV. Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. *Int J Community Med Public Health* 2016;3:1612-6.