

Evaluation of Multiple-Choice Questions by Item Analysis, from an Online Internal Assessment of 6th Semester Medical Students in a Rural Medical College, West Bengal

Sharmistha Bhattacharjee, Abhijit Mukherjee, Kallol Bhandari, Arup Jyoti Rout

Department of Community Medicine, North Bengal Medical College, Darjeeling, West Bengal, India

Abstract

Background: Properly constructed single best-answer multiple choice questions (MCQs) or items assess higher-order cognitive processing of Bloom's taxonomy and accurately discriminate between high and low achievers. However, guidelines for writing good test items are rarely followed, leading to generation and application of faulty MCQs. **Materials and Methods:** During lockdown period in 2020, internal assessment was taken through online mode using Google Forms. There were 60 'single response type' MCQs, each consisting of single stem and four options including one correct answer and three distractors. Each item was analyzed for difficulty index (Dif I), discrimination index (DI), and distractor efficiency (DE). **Results:** The mean of achieved marks was $42.92 \pm$ (standard deviation [SD], 5.07). Dif I, DI, and DE were $47.95 \pm$ (SD 16.39) in percentage, $0.12 \pm$ (SD 0.10), and $18.42 \pm$ (SD 15.35), respectively. 46.67% of the items were easy and 21.66% were of acceptable discrimination. Very weak negative correlation was found between Dif I and DI. Out of total 180 distractors, 51.66% were nonfunctional one. **Conclusion:** Item analysis and storage of MCQs with their indices provides opportunity for an examiner to select MCQs of appropriate difficulty level as per the need of assessment and decide their placement in the question paper.

Keywords: Bloom's taxonomy, difficulty index, discrimination index, distractor efficiency, item analysis, multiple-choice questions

INTRODUCTION

Assessment of students by multiple-choice question (MCQ or item) is an well acceptable method for its (1) objectivity, (2) comparability, and (3) minimized assessor's bias.^[1]

In India, single best-answer MCQs have been commonly used for medical entrance and university examinations.^[2] It is a popular tool of assessment because such tests can be taken for a large number of students, easily scored, help in controlling cheating, and enable teachers to cover a wider range of syllabus. These types of questions were twice more reliable in evaluation of the students' knowledge compared to short-answer questions.^[3] Properly constructed MCQs assess higher-order cognitive processing of Bloom's taxonomy (interpretation, synthesis, and application of knowledge) instead of just testing recall of isolated facts and are thus able to accurately discriminate between high and low achievers.^[4,5]

One best response type MCQs consist of a stem, one correct or best response (key), and few more wrong choices (distractors).^[6]

The main challenge in preparing MCQs is to construct good test items, which requires good depth of knowledge of the subject, understanding of the objectives of assessment, and good skills in writing the items.^[7,8] Obviously, there are many guidelines for writing good test items but they are rarely followed, leading to the generation and application of faulty MCQs.^[9]

Item analysis is a process, which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. It is especially valuable in improving items, which will be used again in later tests.^[10]

Address for correspondence: Dr. Arup Jyoti Rout,
Department of Community Medicine, North Bengal Medical College,
Darjeeling - 734 012, West Bengal, India.
E-mail: dr.arupjyoti81@gmail.com

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Bhattacharjee S, Mukherjee A, Bhandari K, Rout AJ. Evaluation of multiple-choice questions by item analysis, from an online internal assessment of 6th semester medical students in a rural medical college, West Bengal. Indian J Community Med 2022;47:92-5.

Received: 26-08-21, **Accepted:** 12-02-22, **Published:** 16-03-22

Access this article online

Quick Response Code:



Website: www.ijcm.org.in

DOI: 10.4103/ijcm.ijcm_1156_21

Due to countrywide lockdown in 2020 owing to COVID pandemic, conventional internal assessment (offline answering of long and short answer type questions) was not conducted in a rural medical college, West Bengal. Hence, the Department of Community Medicine had decided to take the test in online mode using MCQs, followed by item analysis.

MATERIALS AND METHODS

Ninety-eight MBBS students of 6th semester appeared for an internal assessment on August 14, 2020, through online mode using Google Forms. There were 60 “single response type” MCQs consisting 1 mark each without any negative marking for wrong answer/s. The time allotted was 80 min. The MCQs were constructed by all teachers in the department. All MCQs had single stem, one correct answer (key), and three incorrect alternatives (distractors). Each item was analyzed for difficulty index (Dif I), discrimination index (DI), and distractor efficiency (DE). The data so obtained were entered in MS Excel 2019 and analyzed. Scores of 98 students were arranged in descending order and were divided into three groups. The first group consisting of 1/3rd of total students with higher marks (top third) are labeled as high achievers and the 2nd group consisting of 1/3rd of total students with lower marks (bottom third) are labeled low achievers. Middle 1/3rd was discarded.

Calculations were made using the following formulae:^[11,12]

- $Dif I = (h + l)/n \times 100$
- $DI = 2 (h - l)/n$.

Where;

h = Number of students answering correctly in high achievers' group = 33 students

l = Number of students answering correctly in the low achievers' group = 33 students

n = Total number of students in both groups including nonresponders = 66 students

- Interpretation
- Difficulty Index (Dif I)

Difficulty index describes the percentage of students who answered the item correctly and ranges between 0 and 100%. The higher the Dif I value; the lower is the difficulty (easy) and the lower the Dif I value; the greater is the difficulty of an item. Dif I >70% is considered as easy items, <30% as difficult and in-between percentage are acceptable.

Discrimination index

DI is the ability of an item to distinguish between high and low achievers. It ranges from 0 to ≥ 0.4 . Higher the DI, better the discrimination among high and low achievers.^[13] Negative DI means defective item/wrong key and the students of lower ability answer more correctly than those with higher ability.

Distractor efficiency

Students who have not mastered the subject should choose the distractors more often, whereas the well-prepared

students should discard them more frequently while choosing the correct option. Any distractor that has been selected by <5% of the students is considered to be a non-functional distractor (NFD).^[14] Items containing no NFDs have 100% DE, while items with 3 NFDs have no DE.

RESULTS

Sixty MCQs with their 240 options (60 correct options and 180 distractors) were analyzed. The mean of achieved marks was $42.92 \pm$ (standard deviation [SD] 5.07). Dif I, DI, and DE were $47.95 \pm$ (SD 16.39) in percentage, $0.12 \pm$ (SD 0.10), and $18.42 \pm$ (SD 15.35), respectively [Table 1]. Items that can be categorized as difficult are found to be 15%, whereas 46.67% of the items were easy [Table 2]. Items with poor DI were 70% and 21.66% were of acceptable discrimination. Negative discrimination showed by 6.67% of the items [Table 3]. Very weak negative correlation was found between Dif I and DI [Figure 1]. Out of total 180 distractors, 51.66% were nonfunctional one. 1 NFD and 2 NFDs were found in 35% of items each. 16.67% items had all the three distractors as NFDs, whereas only 13.33% items had no NFD [Table 4].

DISCUSSION

One-best multiple-choice questions

A large portion of curriculum is assessed in a short period of time requiring less effort on behalf of the student, although it takes a lot of effort and time spent by the examiner to make high quality one-best MCQs, as compared to descriptive questions. One-best MCQ is an efficient tool in identifying the strengths and weaknesses in students, as well as providing guidelines to teachers on their educational protocols.^[15]

Difficulty index

Dif I, also called ease index, describes the percentage of students who correctly answered the item. It measures ‘How difficult or easy the questions were?’ Too difficult items

Table 1: Distribution of items according to mean \pm standard deviation of outcome variables (n=60)

Outcome variables	Mean \pm SD
Achieved score	42.92 \pm 5.07
Difficulty index	47.95 \pm 16.39
Discrimination index	0.12 \pm 0.10
Distractor efficiency	18.42 \pm 15.35

SD: Standard deviation

Table 2: Distribution of items according to their difficulty index (n=60)

DIF I (%)	Interpretation	Number of items, n (%)
<30	Difficult	9 (15)
30-70	Acceptable	23 (38.3)
>70	Easy	28 (48.7)

DIF I: Difficulty index

Table 3: Distribution of items according to their discrimination index (n=60)

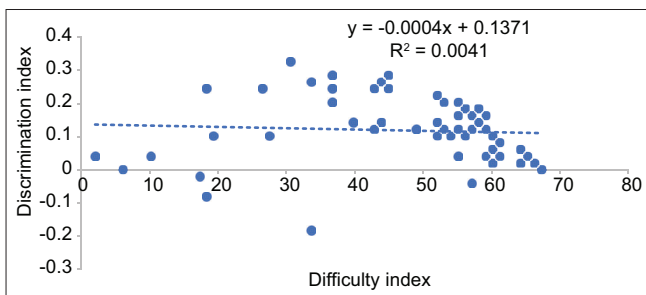
DI	Interpretation	Number of items, n (%)
Negative	Defective item/wrong key	4 (6.67)
0-0.19	Poor discrimination	42 (70)
0.2-0.29	Acceptable discrimination	13 (21.66)
0.3-0.39	Good discrimination	1 (1.67)
≥0.4	Excellent discriminator	0

DI: Discrimination index

Table 4: Distribution of items according to their distractor efficiency (n=60)

NFDs	Distractor efficiency interpretation (%)	Number of items, n (%)
No NFD	100	8 (13.33)
1 NFD	66.66	21 (35)
2 NFD	33.33	21 (35)
3 NFD	0	10 (16.67)

NFDs: Nonfunctioning distractors

**Figure 1:** Distribution of items according to correlation between difficulty index and discrimination index

(DIF I ≤30%) will lead to deflated scores, while the easy items (DIF I >70%) will result into inflated scores and a decline in motivation.^[16]

Two studies had shown that their mean of DIF I were 39.4 ± 21.4 and 52.53 ± 20.59 , respectively.^[1,17] The mean Dif I of the present study was somewhere in between those two findings. The reason behind most of the items being easy could be most of the questions were from ‘must know’ part of the syllabus so proportion of marking the correct option was soaring in both high and low achievers.

Too easy items should be placed either at the start of the test as “warm-up” questions or removed altogether, similarly too difficult items should be reviewed for possible confusing language, areas of controversies, or even an incorrect key.^[18]

Discrimination index

The difficulty and discrimination indices are often reciprocally related. While questions with high Dif I (easier questions) are considered as poor discriminators, questions with low Dif I (harder questions) are considered as good discriminators.^[19] In the present study, most of the items were of poor discrimination.

As we have found that Dif I was mostly easy, assuming that those items were attempted correctly by every student, it renders poor discrimination.

In negative DI, students of lower ability answer questions correctly than those with higher ability. Reasons for negative DI can be wrong key, ambiguous framing of question, or generalized poor preparation of students.^[20] The present study was also not free from wrong key, but the proportion remained below 7%. Another reason may be a student of lower ability by guess selects correct response, while a good student suspicious of an easy question takes harder path to solve and end up being less successful.

Distractor efficiency

It is actually a relationship between the total test score and the distractor chosen by the students.

More nonfunctional distractors (NFDs) 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. The present study showed that more than half of the distractors were NFDs (reduced DE) and most of the test items were easy to answer (increased DIF I). Possible explanation may be inability of the teachers to choose good distractors. However, near similar results were reported by Namdeo and Sahoo with 53.4% NFDs.^[21] However, in contrast, Gajjar *et al.* reported only 11.4% NFDs, while Hingorjo *et al.* reported a mean DE of 81.4%, which is much higher than present mean of DE.^[1,18]

CONCLUSION

MCQs cover wide area of the subject in a short period of time, are preferred method of objective assessment, and selection of good MCQs can obviously judge knowledge of the students. Item analysis is a simple procedure for evaluation of validity and reliability of MCQs. Item analysis and storage of MCQs with their indices provides opportunity for an examiner to select MCQs of appropriate difficulty level as per the need of assessment and decide their placement in the question paper.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- 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.
- Kartha CC. Anguish of a medico. *Curr Sci* 2005;89:725-6.
- Kuechler WL, Simkin MG. How well do multiple choice tests evaluate student understanding in computer programming classes? *J Inf Syst Educ* 2003;14:389-400.
- Carneson J, Delpierre G, Masters K. Designing and Managing MCQs: Appendix C: MCQs and Bloom's Taxonomy; 2011. Available from: <http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html>. [Last

- accessed on 2020 Sep 20].
5. Case SM, Swanson DB. Constructing Written Test Questions for the Basic and Clinical Sciences. 3rd ed. Philadelphia: National Board of Medical Examiners; 2010. Available from: <http://www.nbme.org/publications/item-writing-manual.html>. [Last accessed on 2020 Oct 05].
 6. Haladyna TM. Guidelines for developing MC items. In: Haladyna TM, editor. Developing and Validating Multiple-Choice Test Items. 3rd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004. p. 97-126.
 7. Walsh K. Advice on Writing Multiple Choice Questions (MCQs). *BMJ Careers*; 2005. Available from: <http://careers.bmj.com/careers/advice/view-article.html?id=616>. [Last accessed on 2020 Sep 23].
 8. Huang YM, Trevisan M, Storfer A. The impact of the 'all-of-the-above' option and student ability on multiple choice tests. *Int J Scholarsh Teach Learn* 2007;1:1-13.
 9. Downing SM. The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Adv Health Sci Educ Theory Pract* 2005;10:133-43.
 10. Bai X, Ola A. A tool for performing item analysis to enhance teaching and learning experiences. *Issues Inf Syst* 2017;18:128-36.
 11. Ananthkrishnan N. The item analysis. In: *Medical Education Principles and Practice*. 2nd ed. Pondicherry: JIPMER; 2000. p. 131-7.
 12. Tejinder S, Piyush G, Daljit S. *Principles of Medical Education*. 3rd ed. India: Jaypee Brothers, Medical Publishers Pvt. Limited; 2009. p. 70-7.
 13. Mitra NK, Nagaraja HS, Ponnudurai G, Judson JP. The levels of difficulty and discrimination indices in Type A multiple choice questions of preclinical semester 1 multidisciplinary summative tests. *Jpn Soc Mech Eng* 2009;3:2-7.
 14. 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.
 15. 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 the fellowship of the Royal College of Radiologists in Clinical Oncology. *Clin Oncol (R Coll Radiol)* 2008;20:571-6.
 16. Eaves S, Erford B. The Gale Group. The Purpose of Item Analysis, Item Difficulty, Discrimination Index, Characteristic Curve. Available from: <https://www.education.com/reference/article/itemanalysis>. [Last accessed on 2013 Apr 15].
 17. 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;7:45-50.
 18. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc* 2012;62:142-7.
 19. Carroll RG. Evaluation of vignette-type examination items for testing medical physiology. *Am J Physiol* 1993;264:S11-5.
 20. Scantron Guides – Item Analysis, Adapted from Michigan State University Website and Barbara Gross Devils Tools for Teaching. Available from: <https://www.freepdfdb.com/pdf/itemanalysisscantron>. [Last accessed on 2013 Apr 13].
 21. Namdeo SK, Sahoo B. Item analysis of multiple-choice questions from an assessment of medical students in Bhubaneswar, India. *Int J Res Med Sci* 2016;4:1716-9.