Original Research Article

Impact of distractors in item analysis of multiple choice questions

Ismail Burud[1], Kavitha Nagandla[2]*, Puneet Agarwal[1]

[1]Department of Surgery, [2]Department of Obstetrics and Gyanecology, International Medical University, Jalan Dr Muthu, Bukit Rasah, 70300 Seremban, Negeri Sembilan, Malaysia

*Correspondence:
Dr. Kavitha Nagandla,
E-mail: Kavitha.nagandla@gmail.com

ABSTRACT

Background: Item analysis is a quality assurance of examining the performance of the individual test items that measures the validity and reliability of exams. This study was performed to evaluate the quality of the test items with respect to their performance on difficulty index (DFI), Discriminatory index (DI) and assessment of functional and non-functional distractors (FD and NFD).
Methods: This study was performed on the summative examination undertaken by 113 students. The analyses include 120 one best answers (OBAs) and 360 distractors.
Results: Out of the 360 distractors, 85 distractors were chosen by less than 5% with the distractor efficiency of 23.6%. About 47 (13%) items had no NFDs while 51 (14%), 30 (8.3%), and 4 (1.1%) items contained 1, 2, and 3 NFDs respectively. Majority of the items showed excellent difficulty index (50.4%, n=42) and fair discrimination (37%, n=33). The questions with excellent difficulty index and discriminatory index showed statistical significance with 1NFD and 2 NFD (p=0.03).
Conclusions: The post evaluation of item performance in any exam in one of the quality assurance method of identifying the best performing item for quality question bank. The distractor efficiency gives information on the overall quality of item.

Keywords: Distractor efficiency, Item analysis, One best answer

INTRODUCTION

One best multiple choice questions (MCQ) are an important assessment tool in any examination. Construct of good quality MCQ is time consuming and requires trained faculty. It is important to note that MCQ should not just be testing recall of isolated facts but should assess higher order cognitive process of Blooms Taxonomy such as interpretation, synthesis and application of knowledge.[1,2]

There are two major types of multiple-choice tests, criterion-referenced tests (CRTs) in which the goal is

usually to make a decision about whether or not an individual can demonstrate mastery in an area of content and competencies and norm-referenced tests (NRTs) where the goal is usually to rank the entire set of individuals in order to make comparisons of their performances relative to one another.[3] MCQ consists of a stem with a question followed by number of options. One of the options is the best or correct response known as the key while others are described as distractors.[4] An essential characteristic of distractors is that all options shall present plausible answers and if possible none shall be incorrect.[5] The key to good quality MCQ is based on the availability of good distractors as it is able to

discriminate between the informed and uninformed student.[6]

The concern regarding the construction of MCQ is the reliability of the score and for this reason item analysis is important. Item analysis is a valuable, yet relatively simple procedure performed after the examination that provides information regarding the reliability and validity of a test item.[7] Various terminologies are used in item analysis which needs proper understanding. Difficulty Index (DIFI) reflects the percentage of correct answers to total response and hence tells us how easy or difficult the questions were.[3,8,9] DIFI (p-value) also called ease index ranges from 0-100% and higher the percentage easier is the question.[9] It is calculated by Kelly's Method adding the correctly answered items by the upper 27% and lower 27% of the students' performance divided by the total number of students in both groups.[10] Item difficulty can range from 0.0-1.0(0%-100% and the recommended average level of difficulty should range between 0.31-0.60 (31%-60%).[11]

Discriminatory index (DI) identifies students who are performing well from those performing poorly. It reflects the difference between the percentage of high achieving students who got the answer correct and the percentage of low achieving students who got the answer correct. It is obtained by deducting the correctly responded items in the upper group from the correctly responded students in lower group divided by the number of students in one group.[12] DI is the point-biserial correlation with values ranging from -1 to +1. It is +1 when more students in the upper group (high achievers) answer the item correctly and -1 when the lower achievers answer the item correctly. An Item with a difficulty of 0 or 1 will always have a DI of 0 and DI is maximized when DIFI is close to 0.50. A DI of 0.15- 0.25 is considered desirable.[13]

Distractors are an important components of an item and has a great impact on the total test score. Student's performance depends on how the distractors are designed.[14] For this reason, Distractor efficiency (DE) which indicates whether the distractors in the item are well chosen or have failed to distract students from selecting the right answer is very important. (8) All distractors need to be relatively close to the key of an item. Functional distractors (FD) are those that are selected by >5% or more of the examinees and Non-functional distractors (NFD)are the options selected by <5% of the examines.[11,13,15] DE is determined for each item on the basis of the number of NFDs and ranges from 0-100%.

Objective

To evaluate the One Best Answers or items by assessing the difficulty index (DFI), discriminatory index (DI), Distractor Efficiency (DE). The post evaluation information on items will be utilized to develop question bank following revision /discarding of the items tested.

METHODS

The study was performed in the clinical sciences phase at International Medical University. A total of 113 students who are in the semester 9, year 5 participated in summative examination. The test consists of 120 OBAs based on the assessment blueprint. Each item had four options, one of them being the correct answer and the other three being distractors.

A correct answer was awarded a mark of 1 and 0 for the incorrect answer. The maximum possible score of the test was 90 and minimum is 0. Students' responses from the OBAs were analyzed using Microsoft Excel. The OBAs were analyzed for their level of difficulty, difficulty index (DFI), power of discrimination as measured by the discrimination index (DI), and distractor analysis (DE) for all non-correct options

Item analysis procedure

The test data was entered in the MS excel software and analyzed. The score of 113 students was entered in descending order. Total of 120 OBAs and 360 distractors were analyzed and various indices like DIF I, DI, DE, and nonfunctional distractor (NFD) were calculated with following formulas.

Difficulty index (DFI) =(UG+LG/N) x 100

UG is the upper group of the students with high ability who answered the question correctly and LG is the lower group of the students with low ability who answered the question correctly. N is the total number of the students who answered the question correctly. The higher the difficulty index, the easier the item is understood to be.

Discriminatory Index (DI) =(UG-LG/N)

The item DI is the point biserial correlation between getting the item right and the total score on all other items. Then, the total number of students in the upper 27% who obtained the correct responses and the lower 27% who obtained the correct responses were counted. The higher the DI the better the test item discriminates between the students with higher test scores and those with lower test scores.

Distractor Efficiency (DE) = Is calculated as Non-Functional distractor (NFD) from the distractor that has been selected by less than 5% of the students. The summary of the reference of the item statistics is shown in Table 1 (11).

Statistical analysis

Data is entered in MS Excel 2007 and the item statistics calculated. SPSS version 25 is used to calculate the mean, standard deviation and test of statistical significance by unpaired T test.

Table 1: Reference for item statistics.

Difficulty index (DFI) (Range)	Interpretation	Discriminatory index (DI) (Range)	Interpretation
≤ 30%	Difficult	≤0.19	Poor
31-40%	Good	0.19-0.29	Fair
41-60%	Excellent	0.3-0.39	Good
61-80%	Easy	≥0.40	Excellent
≥81%	Very easy		

Distractor Efficiency (DE)= Distractor efficiency ranged from 0 - 100% and was determined on the basis of the number of NFDs in an item. Three NFD: DE = 0%; 2 NF-D: DE = 33.3%; 1 NF-D :DE =66.6%; No NFD: DE = 100%.

RESULTS

A total of 120 OBAs with 360 distractors were analysed. Out of the 360 distractors, 85 distractors were chosen were chosen by less than 5% with the distractor efficiency of 23.6%. About 47 (13%) items had no NFDs while 51 (14%), 30(8.3%), and 4 (1.1%) items contained 1, 2, and 3 NFDs respectively (Table 2). The DFI, DI and DE were analyzed for each OBA as shown in the Table 2-4. Regarding difficulty index, about 29.6% (n=35), were very easy questions, 15% (n=18), 35 % (n=42) were excellent, 16.6% (n=20), were good and 4.1% (n=5) were very difficult questions.

Table 2: Percentage of non functional distractors.

No. of Non functional distractors	Percentage	Distractor Efficiency (DE)
0NFD	13% (n=47)	100%
1NFD	14% (n=51)	66.6%
2NFD	8.3% (n=30)	33.3%
3NFD	1.1 % (n=4)	0%

Table 3: Item analysis: Difficulty index (DFI) and Discriminatory index (DI).

Difficulty index	Percentage	Discriminatory index	Percentage
Very Easy	29.6% (n=35)	Excellent	20.8% (n=25)
Easy	15% (n=18)	Good	29.1% (n=35)
Excellent	35% (n=42)	Fair	44% (n=37)
Good	16.6% (n=20)	Poor	30.8% (n=13)
Very difficult	4.1% (n=5)	Negative discrimination	8.3% (n=10)

With respect to discriminatory index, (20.8%, n=25) items showed excellent difficulty index, 29.1% (n=35) good discrimination, 44% (n=37) showed fair discrimination, and 30.8% (n=13) and 8.3% (n=10)

showed poor and negative discrimination respectively. The questions with excellent difficulty index and discriminatory index showed statistical significance with 1NFD and 2 NFD (p=0.03). Regarding the Distractor efficiency (DE), about 13 % (n=47) of questions have DE of 100%, 14% (n=51) have DE of 66.6%, 8.3% (n=30) and have DE of 33.3%. About 1.1% (n=4) have DE of 0%.

Table 4: Items statistics relationship between DFI, DI and DE.

Item indicators	Mean score with SD
DFI	62.5 (7.10)
Discriminatory index	0.22 (4.45)

Relationship between the ideal items of good discriminatory index and difficulty index (n=35) is statistically significant (p=0.03)

DISCUSSION

Well-constructed one best MCQ is a good assessment tool which assess the level of cognition (1). In addition to the DIFI and DI, DE is very important as the quality of the distractors has a great impact on the DIFI and DI. Items with NFD (<5% examinees selected the distractor) are important to establish DE. The number of NFDs present in an item range from 0%-100%. DE is indirectly proportional to NFD and items with more functional distractors increase the DE.[9,13] Items with high NFDs reduce both the DE and DI but increases the DIFI; hence the item is easy for the students but a poor discriminator of academic performance.[8] DE is expressed as 0%, 33.3%, 66.6% and 100% depending on number of NFD as 3, 2, 1 and 0 respectively.[12] The selection or rejection of items for question bank is best guided by DE. Items with 0% DE should be discarded whereas those with varying percentages should be revised by replacing the distractors with better choices to be reused in future examinations. It is often necessary to revise items in which the distractor is selected more often than the correct answer.[13]

In present study we found that out that our distractor efficiency is 23.6%. This is lower than the study by Mehta G et al, with 35.3% of NFD and Virendra et al with NFD of 24%.[14,15] Our NFD is higher than Gajjar S et al, with NFD of 11.4%.[16] The present study highlighted a significant relation as observed between distractor efficiency and both item difficulty and discrimination. These results are consistent with previous findings by studies by the authors Tarrant et al, Hingorjo et al.[17,18] There are studies that suggest that numbers of effective distractors are related to discrimination but unrelated to difficulty.

The limitation of the study is the small sample size of 113 students. An increased sample size could improve the evaluation of item and distractor performance. Additionally, we have not analysed the effect of Blooms

level of questions effects on distractor efficiency and discriminatory index. The present findings are encouraging demonstrating that a good item will have a good discrimination index and difficulty index.

CONCLUSION

Performing item analysis is an important aspect of quality assurance of examinations. It is important that the indices of item analysis must be analysed holistically. The decision to revise the questions must be based on good difficulty index, discriminatory index and functional distractors.

REFERENCES

1. Carneson J, Delpierre G, Masters K. Designing and managing MCQs: Appendix C: MCQs and Bloom's taxonomy. 2011. Available at: URL: http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html. Accessed 2 February 2011.
2. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences, National Board of Medical Examiners 3rd ed. 2010. Available at: URL: http://www.nbme.org/publications/item-writing-manual.html. Accessed 5 February 2011.
3. Hotiu A. The relationship between item difficulty and discrimination indices in multiple-choice tests in a Physical science course. Boca Raton, Florida: Florida Atlantic University; 2006. Available at: http://www.physics.fau.edu/research/education/A.Hotiu_thesis.pdf. Accessed 14 June 2015.
4. Cizek GJ, O'Day DM. Further investigation of nonfunctioning options in multiple-choice test items. Educational Psychol Measurement. 1994 Dec;54(4):861-72.
5. Saudi Commission for Health Specialties. Item writing manual for multiple-choice questions. 2015. Available at: http://www.scfhs.org.sa/education/HighEduExams/Guidlines/mcq/Documents/MCQ%20Manual.pdf. Accessed 12 June 2015
6. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education. 2017 Jan 1;4(1):1301013.
7. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005 Jan 1;12(1):19-24.
8. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos University Med J. 2018;18(1):e68-e74.
9. Hingorjo MR1, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012 Feb;62(2):142-7.
10. Linn RL, Miller MD. Measurement and assessment in teaching. Upper Saddle River, New Jersey: Pearson Education Inc. 2005;348-65.11.
11. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non- functioning distractors in multiple choice questions: a descriptive analysis. BMC Med Edu. 2009;9(40):2-8.
12. Ebel RL, Frisbie DA. Essentials of educational measurements. 5th ed. Englewood Clipps, NJ: Prentice Hall; 1991.
13. Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. Natl Med J India. 2008;21:130-3.
14. Gyata Mehta, Varsha Mokhasi. Item Analysis of Multiple Choice Questions- An Assessment of the Assessment Tool. Int J Health Sci Res. 2014;4:197-202.
15. Virendra C Patil, Harsha V Patil. Item Analysis of Medicine Multiple Choice Questions (MCQs) for Under Graduate (3rd year MBBS) Students. Res J Pharmaceut Biol Chem Sci. 2015;6:1242-51.
16. 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 Comm Med. 2014 Jan;39(1):17-20.
17. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Med Edu. 2009 Dec;9(1):40.
18. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pakistan Med Assoc. 2012 Feb 1;62(2):142-7.