Item Analysis: The impact of distractor efficiency on the discrimination power of multiple choice items

Assad Ali Rezigalla
University of Bisha

Elwathiq Khalid Ibrahim (elwathiq@ub.edu.sa)
University of Bisha

Amar Babiker ElHussein
Medicine Program, Nile College, Khartoum, Sudan

Research article

Keywords: item analysis, Distractor efficiency, difficulty index, discrimination index, correlation

Posted Date: October 10th, 2019

DOI: https://doi.org/10.21203/rs.2.15899/v1

Abstract

Background Distractor efficiency of multiple choice item responses is a component of item analysis used by the examiners to to evaluate the credibility and functionality of the distractors.Objective To evaluate the impact of functionality (efficiency) of the distractors on difficulty and discrimination indices.Methods A cross-sectional study in which standard item analysis of an 80-item test consisted of A type MCQs was performed. Correlation and significance of variance among Difficulty index (DIF), discrimination index (DI), and distractor Efficiency (DE) were measured.

Results There is a significant moderate positive correlation between difficulty index and distractor efficiency, which means there is a tendency for high difficulty index go with high distractor efficiency (and vice versa). A weak positive correlation between distractor efficiency and discrimination index.

Conclusions Non-functional distractor can reduce discrimination power of multiple choice questions. More training and effort for construction of plausible options of MCQ items is essential for the validity and reliability of the tests.

Introduction

Well-constructed multiple-choice questions (MCQs) are appropriate tools for the assessment of the cognitive learning domain. It can test a wide range of knowledge, including; recalling, comprehension, and problem-solving, with high objectivity and accurate interpretation of content validity (1).

The effectiveness of MCQs in assessing the learning of the students can be measured both by pre-validation and post-validation methods. Item analysis is a statistical process which is used as a post-validation method for measuring the effectiveness of MCQs regarding their validity and reliability (2). Item analysis is a mathematical analysis of students' responses on an exam (test) to evaluate the quality items and consequently improving the assessment (3). The main advantage of item analysis is the ability to increase the effectiveness of the exam. The effectiveness of the exam is improved either by refining the defected items or deletion of poorly constructed ones from the questions bank (3, 4). Item analysis includes three components; difficulty index (DIF), discrimination index (DI), and distractor Efficiency (DE) (5, 6). The distractor Efficiency (DE) is a component of item analysis that allows the assessor to evaluate the credibility of incorrect options (distractors) of MCQ item. The distractor is considered functioning (FD) if selected by not less than 5% of the group of the examinee (7). Distractors efficiency (DE) is calculated according to the number of non-functional distractors per item (NFD) (8).

This study was conducted to evaluate the impact of non-functional distractors (NFD) on the difficulty and discrimination indices of the test items.

Materials And Methods

Study design

This is a cross-sectional, analytic study. It was conducted at the University of Bisha, College of Medicine (UBCOM) in the duration from April to June 2018. UBCOM adopts a three-phase, integrated approach to undergraduate medical curriculum.

Material

One of the phase I modules exams, the Principles of Human Diseases was selected for the study. This is an integrated, multi-disciplinary module which is conducted during the second semester of year two (number of the students was 45).

The test was consisted of 59 items of multiple choice questions (type A MCQs). Each item is formed of a stem followed by four options, a single best answer, and three distractors. No penalty for the blank or wrong answers.

The exam blueprint was developed by the course instructors and reviewed by A standing Students Assessment Committee (9). Following the exam, standard item analysis was obtained (Apperson, Data Link 1200) and processed for the study.

To calculate the distractors efficiency, the items were classified according to the number of non-functional distractors into; poor (3NFD), moderate (2NFD), good (1NFD) and excellent (0NFD) (3, 8, 10, 11).

Ethical consideration

The study was approved by UBCOM research and ethics committee.

Statistical analyses

The data obtained from the standard item analysis were analyzed by using SPSS V20 (Armonk, NY: IBM Corp, USA). Descriptive statistics and Pearson correlation coefficient were applied to measure the significance of difference and correlation among different variables. Level of significance was fixed at 95%, and any $P < 0.05$ was considered to be significant.

Results

Item analysis:

The total number of items analyzed was 59. The average score of the class was 55.5 (69.38%). Class median was 56.0 (70.0%). Highest and lowest scores were 78 (97.50%) and 35 (43.75%) respectively. KR-20 was 0.906. The average DIF and DIS of the test were 69.4 (±21.86) and 0.3 (±0.16), respectively (Table 1).

Items were classified according to DE, DIS, and DIF table 2, 3, and 4, respectively.

There is a significant moderate positive correlation between DIF and DE, which means there is a tendency for high DIF go with high DE (and vice versa). A weak positive correlation between DE and Disc Index

(Table 5).

	Mean	Median	Minimum	Maximum	Standard Deviation
DIF	37.5	35.48	3.23	87.10	19.046446
DISC	0.46	0.5	0	0.88	0.22972616

Table 1: The descriptive statistics of exam items.

Distractor efficiency (DE)	Frequency	Percent
Poor (3NFD) 25	2	3.4
Moderate (2NFD) 50	13	22.0
Good (1NFD) 75	22	37.3
Excellent (0NFD)100	22	37.3
Total	59	100.0

Table 2: The distractor efficiency of exam items. Classification of exam items according to the number of nonfunctional distractors (NFD).

Discrimination index	Frequency	Percent
Cannot discriminate	2	3.4
Acceptable	6	10.2
Good	10	16.9
Excellent	41	69.5
Total	59	100.0

Table 3: The discrimination index of exam items. Classification of exam items according to the discrimination index of exam items.

difficulty index of the exam (DIF)	Frequency	Percent
Easy	2	3.4
Difficult	14	23.7
Acceptable	43	72.9
Total	59	100.0

Table 4: The difficulty index of exam items. Classification of exam items according to the difficulty index of exam items.

Correlations

		DE	DIF	DIS
DE	Pearson Correlation	1	.538**	.259*
	Sig. (2-tailed)		.000	.047
	N	59	59	59

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 5: The correlation between DE, DIF, and Disc.

Discussion

The number of exam items was adjusted according to the course blueprint and the tested domains.

The KR-20 examination was 0.906, ideal, and showing high reliability of the standard examination (12, 13). Values such as 0.8 and higher are the aims of medical education. This finding is in agreement with the work of Kehoe (12-14). He reported that for short tests (10-15 items) values of as low as 0.5 are reasonable, but those tests with more than 50 items should yield values of 0.8 or higher. Low values of KR-20 were linked to many easy or difficult questions, poorly written nondiscriminating items, non-homogeneity of educational contents, and
the discrepancy between the assessment level and the educational task(14, 15).

The both of majority of exam items (72.9%) and the average exam difficulty (69.4±21.86) were within the acceptable difficulty index range. Moreover, 69.5% of the items were categorized as excellent discriminating, and the average exam discrimination index was 0.3 (±0.16).

The type of correlation between DE and DIF indicates that items with less non-functional distractors have high difficulty index (easy items) and vice versa. This finding in agreement with recent works of Hingorjo et al., Burud et al. and Kheyami et al. (8, 16, 17). The decreased number of non-functional distractor increases the difficulty index of items (become easier). Also, they reported that the DE reduces the DIS of items. In the current study, DE has a weak positive correlation with the DIS (P=0.047437, is significant at $p < .05$). NFDs can affect the discrimination power of the item (8) and should be replaced by more plausible distractors (11) or the item removed from the test (18). Such items have high DIF (8) as all students well got them right 23(11)or become distracting and causing a false assessment (10).

NFDs were linked to minimal training of items writing and distractors selection (7, 19, 20). It is clear that DE has an impact on both DIF and DIS individually, but whether it can affect both of them at the same time, need more research work.

Items with nonfunctional distractors can be present in any exam or test; the second step after defining them in the running exam remains open. In such items, the nonfunctional distractors can be changed with more plausible ones or deletion of the question from the bank. The area of debate is the status of these items in the current exam or test. In the current study deletion of items with two or three non-functional distractors increased the average difficulty index of the exam to from 36.83 to 42.82 and DE showed non-significant correlation with DIF of items (r= 0.2296, p= 0.133806). Deletion of such items from exam or test can affect students results and raises ethical debate. Kehoe (1995) reported that deletion of such items is ethical and justifiable (14). He argued that the test
aims to determine the rank of each student. Using items or questions with unacceptable psychometrics is against this objective, and the accuracy of the resulting ranking is degraded.

Limitations of this study include the fewer number of students and items and application on one course. The strength of the study, the test is considered valid and reliable.

Conclusions

Non-functional distractor can reduce discrimination power of multiple choice question. More training and effort for construction of plausible options of MCQ items is essential for the validity and relaibilty of the tests.

Abbreviations

DE: Distractor Efficiency

DIF: Difficulty Index

DI: Discrimination Index

UBCOM: University of Bisha, College of Medicine

Declarations

Confilct of Interest:

None of the authors has a financial or professional benefit that affect the scientific judgement of the study.

Funding:

The study received no funding

Availability of Data:

The dataset which is used and analyzed to derive the results of this article are available and ready to be provided by the corresponding author on reasonable request.

Consent for Publication:

Not applicable

Ethical Approval:

The study was approved by the Research and Ethics committee of theUniversity of Bisha, College of Medicine.

Authors Participation:

All authors made substantial contributions to the conception and design of the study and to the analysis of the data. RAA did the data collection and entry and drafted the literature and methodology section of the article. IEK revised the article and added the discussion and conclusion section. EAB performed data analysis and reporting.

Acknowledgements:

The authors wish to express their acknowlegent to Dr. Ali el-eragi, Associate Professor of Microbiology for providing the rough data (the exm papers, blueprint and item analysis documents).

References

1. Sahoo DP, Singh R. Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. International Journal of Research in Medical Sciences. 2017;5(12):5351.
2. Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. International Journal of Educational and Psychological Researches. 2016;2(4):201.
3. 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. Journal of Health Specialties. 2014;2(4):148.
4. 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(1):19-24.
5. Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S, et al. Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University. 2018;13(4):310-5.
6. Tavakol M, Dennick R. Post-examination analysis of objective tests. Medical Teacher. 2011;33(6):447-58.
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;9:40.
8. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. JPMA-Journal of the Pakistan Medical Association. 2012;62(2):142.

9. Abdellatif H, Al-Shahrani AM. Effect of blueprinting methods on test difficulty, discrimination, and reliability indices: cross-sectional study in an integrated learning program. Advances in medical education and practice. 2019;10:23.
10. 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: official publication of Indian Association of Preventive & Social Medicine. 2014;39(1):17.
11. 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):1.
12. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. Bmj. 1997;314(7080):572.
13. Carmines EG, Zeller RA. Reliability and validity assessment: Sage publications; 1979.
14. Kehoe J. Basic item analysis for multiple-choice tests. Practical assessment, research & evaluation. 1995;4(10):20-4.
15. van de Watering G, van der Rijt J. Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. Educational Research Review. 2006;1(2):133-47.
16. Burud I, Nagandla K, Agarwal P. Impact of distractors in item analysis of multiple choice questions. International Journal of Research in Medical Sciences. 2019;7(4):1136.
17. 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 Medical Journal. 2018;18(1):e68.
18. Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. Applied Measurement in Education. 1989;2(1):51-78.
19. Schuwirth LW, Van Der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? Medical education. 2004;38(9):974-9.
20. Crehan KD, Haladyna TM, Brewer BW. Use of an inclusive option and the optimal number of options for multiple-choice items. Educational and Psychological Measurement. 1993;53(1):241-7.