



## An Item Analysis of Written Multiple-Choice Questions: Kashan University of Medical Sciences

Mansour Sayyah<sup>1</sup>, Zarichehr Vakili<sup>1</sup>, Negin Masoudi Alavi<sup>2</sup>, Manzardokht Bigdeli<sup>1</sup>, Alireza Soleymani<sup>1</sup>, Mohammad Assarian<sup>3</sup>, Zohreh Azarbad<sup>4\*</sup>

<sup>1</sup> Education Developing Center, Kashan University of Medical Science, Kashan, IR Iran

<sup>2</sup> Truma Nursing Research Center, Faculty of Nursing and Midwifery, Kashan University of Medical Science, Kashan, IR Iran

<sup>3</sup> Department of Computer Engineering, Azad University of Kashan, Kashan, IR Iran

<sup>4</sup> Department of Social Medicine, Kashan University of Medical Sciences, Kashan, IR Iran

### ARTICLE INFO

**Article type:**  
Research Article

**Article history:**  
Received: 23 Oct 2012  
Revised: 01 Nov 2012  
Accepted: 07 Nov 2012

**Keywords:**  
Education, Nursing, Baccalaureate  
Reference Standards  
Instrumentation

### ABSTRACT

**Background:** Multiple Choice Questions (MCQs) are one of the most common types of exams used in evaluation of students in any educational setting. The question items making up these exams need to be examined if they are to meaningfully contribute to the student scores. Such characteristics are amenable to examination by item analysis.

**Objectives:** The purpose of this research was to examine the quality of MCQs used in Nursing and Midwifery Faculty and to compare the results with the other faculties in Kashan University of Medical Science in the academic year 2008-2009.

**Materials and Methods:** In this cross-sectional study, 101 multiple-choice exams were randomly chosen for the study, and 37 exams were selected from the faculty of nursing and midwifery. The difficulty, discrimination indices and Cronbach's Alpha were calculated for every exam and then mean values for each index were calculated by LERTAL 5.0 software purchased from Assessment Systems Corporation of the United States.

**Results:** A total of 7062 MCQs in the university and 1793 items in the faculty of nursing and midwifery presented to the students by different instructors were analyzed. The average of difficulty index of the faculty of nursing was 0.5. The discrimination index was 0.36, and the average of Alpha-Cronbach was 0.82 in the faculty of nursing. All the values were significantly better in the faculty of nursing and midwifery compared to the rest of the university.

**Conclusions:** The difficulty index, the discrimination index and the Alpha-Cronbach values in the faculty of nursing were within the acceptable range recommended by experts in the field of educational measurement. However, some of the tests had values less than the recommended.

Published by Kowsar Corp., 2012. cc 3.0.

#### ► Implication for health policy/practice/research/medical education:

This article might be helpful for nursing educators and clarify the importance of item analysis in multiple choice examinations.

#### ► Please cite this paper as:

Sayyah M, Vakili Z, Masoudi N, Bigdeli M, Soleymani A, Assarian M, et al. An Item Analysis of Written Multiple-Choice Questions: Kashan University of Medical Sciences. *Nurs Midwifery Stud.* 2012;1(2):83-7. DOI: 10.5812/nms.8738

\* Corresponding author: Zohreh Azarbad, Department of Social Medicine, Kashan University of Medical Science, Ghotb Ravandi Highway, Kashan, IR Iran. Tel: +98-3615550021, Fax: +98-3615556633, E-mail: azarbad\_zo@yahoo.com

DOI: 10.5812/nms.8738

© 2012 Kashan University of Medical Sciences; Published by Kowsar Corp.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Background

The evaluation of knowledge is understood as an essential component of nursing education. In the evaluation of knowledge levels, different approaches including paper-and-pencil tests, written assignments, oral presentations, and portfolios are understood as a strategy within nursing education programs. A common written test format used across all nursing education settings are Multiple Choice Questions (MCQs) (1, 2). These are the most common types of tests used by the majority of educational institutions, including universities (1). MCQs have been consistently criticized for having several weaknesses, such as decreased validity due to guessing and failure to credit knowledge (3) which can also have negative effects on students' knowledge. The reason is that MCQs expose students to incorrect answers (4). There are many reasons why instructors like these types of questions. Perhaps foremost among them is the fact that such tests can be easily scored, these tests can also help control cheating, and enable instructors to ask questions that cover a wider range of material (5). High quality MCQs are difficult to construct but are easily and reliably scored (6). The well-prepared test items also require students to use a higher level of cognitive processing, which is an advantage of MCQs(7). A study showed that these types of questions were twice more reliable in evaluation of the students' knowledge compared to short-answer questions (5). Although another study concluded that in evaluation of the students' ability to perform in clinical situations, the short-answer format examinations provide a better measure compared to MCQs(8). The questions making up these types of exams need to possess certain psychometric properties if they are to be considered as a reliable instrument. These types of tests are amenable to various types of evaluation by computer software in order to determine their psychometric properties. Item analysis is a procedure to check the properties of every item used in a question (9). Item analysis is widely used to improve test quality through knowledge about item statistics. It allows us to observe the characteristics of a particular item and can be used to ensure that questions are of an appropriate standard for inclusion in a test. Typically, in analysis of a test, two values are computed, a difficulty level and a discrimination index (10). While the difficulty index refers to the difficulty of an item for the respondents to identify the correct alternative among the various choices, discrimination index indicates how well the item discriminate the strong students from the weak ones, and the internal consistency demonstrates the consistency of response among the items measuring a concept (11). A study showed that 18% of items in MCQs were rejected either due to both difficulty level and discrimination index (10). Another study showed that in the University of Ontario, the mean item discrimination coefficient of MCQs was +0.25, with more than 30% of items having unsatisfactory coefficients less than +0.20 and

45% of distracters were non-functioning (12). Are these results due to MC tests lacking the basic necessary properties? There are rich sources of references in regard to the significance of these concepts as well as the acceptable values for these indices (13, 14). Item difficulty within the range of 0.30 to 0.70 is considered as an acceptable index for multiple-choice exams (14). The internal consistency criteria known as the Cronbach alpha is another index that is used to judge a question. Burch (2008) claims that it is necessary to determine reliability of a test for issuing the certificate of competency for medical practice (13). When designing MCQs, the distracters offered to the test takers are also important. Placing a distracter or distracter that is chosen by none of the take testers reduces the number of alternatives and increases the likelihood of guessing an item correctly (15).

## 2. Objectives

Considering the importance of such a criteria in designing MCQs, this descriptive research was designed to determine the item difficulty, item discrimination, internal consistency and distracters used in final examinations of faculty of nursing and midwifery and to compare the results with the values of the Kashan University of Medical Sciences in the academic year 2008-2009.

## 3. Materials and Methods

This cross-sectional research was performed on the 101 randomly selected MCQs in the Kashan University of Medical Sciences. The 37 tests were from the faculty of nursing and midwifery. Item analysis was employed on the items by Laboratory of Educational Research Test Analysis Package (LERTAP) version 5.0d of these exams, including item difficulty, item discrimination, Cronbach alpha and frequencies of correct responses as well as the distracters calculated by the software were transferred to SPSS version12 for further analysis. The difficulties index categories were set to less than 0.30, 0.30 to 0.70 and above 0.70. The discrimination index was classified into five categories to zero, more than zero to 0.20, 0.21 to 0.40, and 0.41 to 0.80 and over 0.81, respectively. The frequency of Alpha-Cronbach index for the entire test was classified into five categories, including 0 to 0.20, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80 and 0.81 and higher. The tests and their designers were kept anonymous and the ethical committee in the Kashan University of Medical Sciences approved the study.

## 4. Results

Overall, 1793 MCQs in 37 exams in the faculty of nursing and midwifery and 7062 items in 101 exams in other faculties in different subjects given by different instructors were analyzed (Table 1). Table 1 shows that 17.7 percent of exams in the faculty of nursing and midwifery had item difficulty less than 0.30 the rate was 21.7 in the university.

**Table 1.** Frequency Distribution of Classified Difficulty Index

Range of Difficulty Index	Faculty of Nursing and Midwifery, No (%)	The University, No (%)
Less than 0.30	318 (17.7%)	1533 (21.7%)
0.30-0.70	1011 (56.4%)	3080 (43.6%)
0.71-1	464 (25.9%)	2449 (34.7%)
total	1793 (100%)	7062 (100%)
P value	0.0001	

**Table 2.** Frequency Distribution of Classified Discrimination Index

Range of Discrimination Index	Faculty of Nursing and Midwifery, No (%)	The University, No (%)
Negative to zero	253 (14.1%)	1258 (17.8%)
0.0 to 0.20	199 (11.1%)	1823 (25.8%)
0.21-0.4	359 (20.0%)	1542 (21.8%)
0.41-0.80	699 (39.0%)	2105 (29.9%)
0.81-1	283 (15.8%)	334 (4.7%)
Total	1793 (100%)	7062 (100%)
P value	0.0001	

**Table 3.** Frequency Distribution Table of Classified Cronbach Values

Range of Cronbach's Alfa Values	Faculty of Nursing and Midwifery, No (%)	The University, No (%)
Less than 0.20	1 (2.7%)	10 (6.9%)
0.20-0.40	2 (5.4%)	5 (3.4%)
0.41-0.60	3 (8.1%)	22 (15.2%)
0.61-0.80	4 (12.8%)	26 (17.9%)
0.81-1	27 (72.3%)	38 (56.6%)
total	37 (100%)	101 (100%)
P value	0.005	

**Table 4.** Frequency Distribution of Selected Distracters

Selected Distracters	Frequency, (%)
All	346, (19.3)
1-non-selected	643, (35.9)
2-non-selected	504, (28.1)
3-non-selected	255, (14.2)
4-non-selected	45, (2.5)
Total	1793, (100)

The 25.9 percent of the exams had item's difficulty over 0.70 in nursing and midwifery faculty; the percentage was 34.7 in the university. The difference was significant (Chi square = 95, P value = 0.0001). The discrimination index for the items in the faculty of nursing and midwifery with negative or zero were 14.1% the percent was 17.8 in the university. The difference was significant (Chi square = 438, P value = 0.00001) (Table 2). Table 3 revealed that 2.7% of the questions in the faculty of nursing and midwifery had an internal consistency less than 0.20 and the percentage was 6.9 in the university. The 72.3% of MCQs

showed a consistency index over 0.81 or more in the faculty of nursing, the percentage was 56.6 in the university. The difference was significant (Chi square = 14.5, P value = 0.005). Finally, the distracters analysis showed that 19.3 percent of items contained all distracters that were sufficiently distracting to be selected by some respondents, while 35.9 of the items had one, 28.1% had two, 14.2% had 3 and 2.5% had four unselected choices (Table 4).

## 5. Discussion

The results showed that the difficulty index, discrimination index and the Alpha-Cronbach values in the faculty of nursing and midwifery were within the acceptable range recommended by experts in the field of educational measurement. The measured indexes were significantly better in the faculty of nursing compared to the rest of the university. Some of the questions that was evaluated in the study had insufficient psychometrics property to be included in the exam. Under such circumstances, tests may lead to the incorrect evaluation of students (16). Results of this research showed that the average of item difficulty for the test conducted at the Nursing, and Midwifery Faculty was 0.54. This value is approximately close

to what Gronlund in 1985 recommended and is with the range 0.3 to 0.70 that Nelson in 2001 suggested (14, 17). However, 25.9 percent of tests items showed item difficulties over the 0.70 criterion. This condition indicates that some of the test items were relatively difficult. When an item difficulty approaches the high value such as some of the items identified in this research, it indicates that either the instructor did not cover the subject matter thoroughly or the student did not show enough interest to study them well (11). In the present research, the average of discrimination index was 0.36. This value is with the range that has been suggested by other investigators (18). In a study in dental college in Pakistan, the 62 percent of items had excellent Discrimination Index (15). However, 14.1 percent of items showed negative discrimination values or values close to zero. Such items are not discriminating the good students from the weak ones plus do not accounts for the true total test variance (11). These items need complete revisions. The value of internal consistency may change by eliminating test items with the low coefficient (19, 20). Finally, the distracter analysis revealed that only 19.3 % of the all the distracters were sufficiently attractive to be selected. Such property of the distracters implies that not all the distracters are fulfilling the objective of the test constructor. Moreover, a study showed that the properties of three stem MCQs were comparable with the four- one (21, 22). It seems that the use of three stem MCQs might be better than four ones with non-functioning distracters. In summary, the results of item analysis of MCQs used in the nursing and midwifery faculty indicated the fact that considerable test items passed the criterion recommended by experts in the field. However, some test items were not well prepared. The quality of measured indexes was better in the faculty of nursing, and midwifery compared to the rest of the university. The reason might be that faculty members in nursing and midwifery school pass several courses in educational subjects, including the MCQs preparation. We recommend that the educational courses add to the programs of the potential faculty members in master and PhD degrees. Further research and reevaluation of the questions may lead to improvement in test constructions by the instructors at this faculty. MCQs have been used extensively in nursing as an evaluation tool in both pre and post registration educational contexts. Our findings indicate that there is considerable room for improvement in the quality of MCQs. We suggest that instructors consider improving the quality of their MCQs by conducting an item analysis and by modifying distracters that impair the discriminatory power of items. The current research provides some data about the very basic characteristics of the exams and its items. The further studies need to be implemented to evaluate the items according to the Blooms Taxonomy of learning domains.

## Acknowledgments

This research project has been granted and approved by deputy of research in Kashan University of Medical Sciences.

## Authors' Contribution

All the authors contributed equally in data gathering and data analysis. The first author designed the research and the corresponding author prepared the article.

## Financial Disclosure

None declared.

## Funding/Support

None declared.

## References

1. Bailey PH, Mossey S, Moroso S, Cloutier JD, Love A. Implications of multiple-choice testing in nursing education. *Nurse Educ Today*. 2012;**32**(6):e40-4.
2. Collins J. Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics*. 2006;**26**(2):543-51.
3. Lau PNK, Lau SH, Hong KS, Usop H. Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests. *Educ Tech Soc*. 2011;**14**(4):99-110.
4. Fazio LK, Agarwal PK, Marsh EJ, Roediger HL, 3rd. Memorial consequences of multiple-choice testing on immediate and delayed tests. *Mem Cognit*. 2010;**38**(4):407-18.
5. Kuechler WL, Simkin MG. How well do multiple choice tests evaluate student understanding in computer programming classes? *J Inf Syst Educ*. 2003;**14**(4):389-400.
6. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas*. 2004;**64**(3):391-418.
7. Clifton SL, Schriener CL. Assessing the quality of multiple-choice test items. *Nurse Educ*. 2010;**35**(1):12-6.
8. Prihoda TJ, Pinckard RN, McMahan CA, Jones AC. Correcting for guessing increases validity in multiple-choice examinations in an oral and maxillofacial pathology course. *J Dent Educ*. 2006;**70**(4):378-86.
9. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ*. 2004;**38**(9):974-9.
10. Zaman A, Niwaz A, Faize FA, Dahar MA. Analysis of Multiple Choice Items and the Effect of Items' Sequencing on Difficulty Level in the Test of Mathematics. *Eur J of Soc Sci*. 2010;**17**(1):61-7.
11. Mitra N, Nagaraja H, Ponnudurai G, Judson J. The Levels Of Difficulty And Discrimination Indices In Type A Multiple Choice Questions Of Pre-clinical Semester 1, Multidisciplinary Summative Tests. *Int e-J Sci Med Educ*. 2009;**3**(1):2-7.
12. DiBattista D, Kurzawa L. Examination of the Quality of Multiple-choice Items on Classroom Tests. *Can J Scholarship Teach Learn*. 2011;**2**(2):4.
13. Burch VC, Norman GR, Schmidt HG, van der Vleuten CP. Are specialist certification examinations a reliable measure of physician competence? *Adv Health Sci Educ Theory Pract*. 2008;**13**(4):521-33.
14. Nelson LR. Item Analysis for Tests and Surveys Using Lertap 5. Perth, Western Australia: Curtin University of Technology; 2000.
15. 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.
16. Nicol D. E-assessment by design: using multiple-choice tests to good effect. *J Further High Educ.* 2007;**31**(1):53-64.
  17. Gronlund NE, Linn RL. Measurement and evaluation in teaching. New York: *Macmillan*; 1990.
  18. 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.
  19. Cizek GJ, O'Day DM. Further investigation of nonfunctioning options in multiple-choice test items. *Educ Psychol Meas.* 1994;**54**(4):861-72.
  20. Shizuka T, Takeuchi O, Yashima T, Yoshizawa K. A comparison of three-and four-option English tests for university entrance selection purposes in Japan. *Lang Test.* 2006;**23**(1):35-57.
  21. Omirin M. Difficulty and discriminating indices of three multiple-choice tests using the confidence scoring procedure. *Educ Res Rev.* 2007;**1**(2):14-7.
  22. Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educ Meas: Issues Pract.* 2005;**24**(2):3-13.