

# Computer Software Application in Item Analysis of Exams in a College of Medicine

<sup>1</sup>Mohammad Resa Sharif, <sup>2</sup>Seyed Mohsen Rahimi, <sup>3</sup>Mehdi Rajabi, <sup>4</sup>Mansour Sayyah

<sup>1</sup>Assoc. Prof., Department of Pediatrics, Kashan University of Medical Sciences, Kashan, I.R. Iran

<sup>2</sup>Assist Prof., Department of Sport Medicine, Iran University of Medical Sciences, Tehran, I.R. Iran

<sup>3</sup>Assist Prof., Department of Medicine, Kashan University of Medical Sciences, Kashan, I.R. Iran

<sup>4</sup>Department of Medicine, Kashan University of Medical Sciences, Kashan, I.R. Iran

<sup>1</sup>[Dr.sharifmd@yahoo.com](mailto:Dr.sharifmd@yahoo.com), <sup>2</sup>[rahimiseyed@yahoo.com](mailto:rahimiseyed@yahoo.com), <sup>3</sup>[mrajabi@yahoo.com](mailto:mrajabi@yahoo.com), <sup>4</sup>[mansorsayyah@yahoo.com](mailto:mansorsayyah@yahoo.com)

## ABSTRACT

MCQ exams need to be item analyzed to assess students. Computer software is used to perform this task reliably. This cross sectional study was performed on 40 MCQ exams including 2445 multiple choice question items using a software. The difficult and discrimination index in addition to Cronbach's alpha was calculated for every questions and exams by LERTAP 5.0 software from ASC-USA. In addition, mean value for each index for the tests was calculated by SPSS: Pc16.0. The result of statistical analysis revealed that the mean of item difficulty and discrimination index of the exams used at the college of medicine were 0.49 and 0.19, respectively. In addition, the mean of coefficient alpha for the entire tests was 0.74. Only in 12 percent of the questions all the distracters were distracting. The mean of difficulty and discrimination index of MCQ exams of college of medicine was 0.68 and 0.19. These values were close to the recommended values suggested by experts in the field, however, a considerable frequency of these indexes by some tests were low. Further research is needed to construct better tests.

**Keywords:** *Item Analysis, Difficulty index, Discrimination index, Cronbach alpha, Multiple choice exams, Software*

## 1. INTRODUCTION

The first step in evaluating the academic achievement of learners in cognitive domain in an academic system is using achievement tests. The selection of a test that can withstand psychometric evaluation is not an easy task. Computer software is used to perform this task reliably. Despite the fact that this is a routine process in nearly all educational institution, hardly if any attempt has been devoted to carefully examine the contents of a test from this prospective. With the advancement of knowledge in test construction and emergence of computer software to ease this process, it is currently possible to perform through psychometric assessment on cognitively oriented test by computer software. Test with multiple choice questions (MCQ) and analyzing their distracters have become the choice of many examiners for their evaluation purpose. In the field of medicine, MCQ test for assessment of medical undergraduate student provided insight about the learning and competencies of the students and multiple choice questions (MCQs) or "items" became the choice for such assessments. "Item analysis" examines student responses to individual test items (MCQs) in order to assess the quality of those items and test as a whole (1-4). There are numerous research results that show the importance of analyzing MCQ tests. While Haladyna examined the validity of taxonomy of MCQ tests and the writing rules for them (5). Researchers including Shiefflett, Phibbs, & Sage; Farely and McCoubrei studied this type of tests from different prospective (6,7,8). Vyas and Supe conducted a research in regard to the response items to a MCQ test and concluded that three choice responses are more appropriate than the 4 or 5 choices (9). Gajjar, Sharma, Kumar & Rana examined the quality of MCQ tests and concluded that quality MCQs truly assess the knowledge and were able to differentiate the students

of different abilities in a correct manner (10). Item analysis and evaluating the quality of a MCQ test including the number of choices as the distracters is important subject for many researches and research in this area continues (11-17).

In a MCQ test there are indices that determine the difficulty of an item from the prospect of respondent called item difficulty index. This concept refers to the difficulty of an item for the respondents to correctly identify the correct alternative among the various choices. In addition, there is the discrimination index that indicates how well the item discriminate the strong students from the weak ones and the internal consistency demonstrate the consistency of response among the items measuring a concept. There are rich sources of references in regard to the significance of these concepts as well as the acceptable values. For instance, Linn & Gronlund, suggest item difficulty within range from 0.60 to 0.70 as an acceptable index for multiple choice exams (18). High quality multiple choice items are difficult to construct but are easily and reliably scored. One of the issues in determining the desirable difficulty index is the cutoff point for this index. The range for this index varies depending on the view of researchers. Linn & Gronlund claimed that the value between 0.20 to 0.40 are an acceptable values (18). The internal consistency criterion knows as the Cronbach alpha is another index that is used to judge a multiple choice test (19). In this regard, different level for different test purposes has been offered. Linn states that the value for the internal consistency should be between 0.60 to 0.85 (18).

Considering the importance of such criterion in designing MCQ exams and the fact that most of the

researches performed in this regard perform item analysis on one single test item, this descriptive research was designed to determine the item difficulty, item discrimination, internal consistency and distracters used in 40 final examinations of College of Medicine of Kashan University of Medical Sciences in an education year.

## 2. MATERIAL AND METHODS

This was a cross sectional research and 40 tests that employed multiple choice question (MCQ) used by the basic science and clinical examination of medical students were item analyzed by Laboratory of Educational Research1 Test Analysis Package (LRTAP version 5.0). Every exam was item analyzed separately by the software and then the results of analysis of every test including the difficulty and discrimination index in addition to Cronbach alpha and frequencies of correct responses as well as the distracters were determined by the software and transferred to SPSS: pc software for further analysis. Descriptive statistics and chi-squared test was employed to present the results.

## 3. RESULTS

Overall, 2445 multiple choice exams for different subjects given by 40 instructors were analyzed. Descriptive statistics including mean, standard deviation and other indices are presented in table 1.

The difficulties index categories were set to less than 0.30, as difficult, 0.30 to 0.70 as desirable and above 0.70 as easy. Based on this classification, the result is presented in table 2.

Table 2 shows that 33.2 percent of exams had item difficulty less than 0.30, 34.1 percent had difficulty index within the recommended range that is, 0.30 to 0.70 and 32.7 percent of the exams had items difficulty above 0.70. Chi-squared indicated that there was no significant difference among the proportions of the item difficulty levels ( $p < 0.05$ ).

Similar procedure was used to present the discrimination index of the exam. The index was classified into five categories. The base for categorization was negative to zero, more than zero to 0.20, 0.21 to 0.40, and 0.41 to 0.80 and over 0.81. The result of this analysis is presented in table 3.

Table 3 shows that the discrimination index for the items with negative or zero were 21.3% , between 0 to 0.20 were 31.8% , between 0.21 to 0.40 were 20.3%, between 0.41 to 0.80 were 24.5% and above 0.81 to 1 were 2%.

The third index calculated for the 40 tests was Cronbach alpha. The average of this index was 0.60. The frequency of this index for the entire test was classified into 5 categories as 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 result of this classification is presented in table 4.

Table 4 shows that 10 % of the exams have Cronbach's alpha less than 0.20 and 65% have values above 0.81.

Finally, the distracters used in this study were analyzed. The result showed that 12percent of items, all distracters were sufficiently distracting and were selected by some respondents, while 34.6% of the items had one, 38.1% had two and 15.3% had three nonelected choices. The result of this analysis is presented in table 5.

## 4. DISCUSSION

This study was designed to perform an item analysis of a MCQ test items used in a college of medicine. The result indicated that while a considerable number of tests possessed acceptable values proposed by the experts in the field of test construction, some tests were lacking this psychometric property. For instance, the result showed that the average of item difficulty for the test conducted at the college of medicine was 0.49. This value is close to what Lin & Gronlund recommends, that is, with the range 0.3 to 0.70(18). When an item difficulty approaches high value such as those found in this research, it may be due to the intention of the test taker who tries to give difficult items to have a more comprehensive evaluation of the students. In other hand, such condition may arise since the test item is not explained clearly and the learners have a fuzzy understanding about the concept, or probably was an item that was not covered during the course of instruction. In addition, having test items that lie within the range of easy may have different interpretation. For instance, in any subject matter, there is basic concept that the learners definitely must learn them. It is very difficult to tease out these items from those that are redundant concept and every learner can easily answer them. Such items make the distracter options hard to present to the responders. The other index that was evaluated in this research was the discrimination index. The result showed that mean of discrimination index was 0.19. However, 21.3 percent of items used in the exams used at the college of medicine had negative discrimination values. Such items have serious flow in their structures and do not serve as distracter, thus, they need to be revised. Gajjar, Sharma, Kumar & , Rana believe that there may be different reasons for these types of question items (10). It is likely that the negative discrimination item was wrongly keyed, ambiguously framed (12) or poor preparation of students for the exam. Items with negative DI are not only useless, but actually serve to decrease the validity of the test. Revising these items may contribute to the validity of the test on next test (17).

Another index that was examined in this study was the coefficient of alpha or Cronbach's alpha (19). The mean of this index was 0.74. This value is within the range proposed by some experts. Linn suggests the value range 0.60 to 0.85 as a desirable range (18).

Finally, the distracter analysis revealed that 12% of the all the distracters were sufficiently attractive to be selected whereas 34.6 had 1, 38.1 had two and 15.3% had three nonelected distracters. In other words, 34.6 of the questions were three-choices, 38.1 were two-choices and 15.3% were one-choice questions, respectively.

Preparing MCQ test with desirable item content is possible through item analysis. This process if performed consistently and carefully can provide sufficient number of test items based on which a data bank of MCQ items can be established. An item analysis of MCQ test within for a particular subject should not be limited to a single testing approach instead repeated item analysis of tests will finally provide sufficient number of items that possess acceptable item difficulty and discrimination index. This in turn will make up the question items for the question bank. Considering the fact that the advancement of techniques and knowledge about the MCQ test construction is of great concern in all domains of education, finding new methods and techniques that facilitate this process will open the horizon to more quick, reliable and accessible mean of MCQ test construction. Such progress will improve ontology-based MCQ item generation that is still in its infancy will advance. Advancements in Semantic Web technologies have created an interest among researchers in developing ontology-based applications in numerous research areas including the field of question generation that has led to the emergence of ontology-based multiple choice question (MCQ) generation. MCQ items have proved to be an efficient tool for measuring the achievement of learners) 17)

In summary, the results of item analysis of multiple choice tests used in the college of medicine indicated that a considerable test items were within the recommended values by experts. However, some tests did not meet the requirement of a well –designed question

items. One of the strength of this research was the number of different tests for assessing different subjects by different instructors. Most of researches performed in this regard examined one test and performed item analysis for that particular test. This research examined 40 tests with 2445 test items. However, more comprehensive item analysis may examine every test items not only from the psychometric point of view, but examine the taxonomy of the items as well. Further research is necessary to provide test items with sufficient difficulty and discrimination index to build a questions bank.

## ACKNOWLEDGEMENT

The researcher wishes to express his appreciation to the deputy of education as well as the deputy of research of Kashan University of Medical Sciences and personnel of educational development center of research of Kashan University of Medical Sciences for their contribution in completion of this research project. Special thanks to Dr. Esmail Fakharian, the dean of College of Medicine of Kashan University of Medical Sciences, for his assistance and support of this project.

## FINANCIAL DISCLOSURE

The Authors certified that all their affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript were completely disclosed. In addition, none of the authors had any relevant financial interests related to the material in the manuscript.

This work was a research grant no 8652 provided by the undersecretary of Research of Kashan University of Medical Sciences

**Table 1:** descriptive statistic of Item difficulty and Item discrimination of 40 exams

Index	Mean	Standard deviation	Max	Min	Number
Item difficulty	0.49	0.31	1	0	2445
Item discrimination	0.19	0.30	0.94	-0.91	2445
Cronbach's alpha	0.74	0.29	0.98	0	40

**Table 2:** frequency distribution of classified difficulty index

level	Frequency	Percent	Cumulative Percent
easy	812	33.2	33.2
desirable	834	34.1	67.3
difficult	799	32.7	100.0
Total	2445	100.0	-

<http://www.ejournalofscience.org>**Table 3:** Frequency distribution of classified discrimination index

Discrimination Index	Frequency	Percent	Cumulative Percent
negative	522	21.3	21.3
0-0.20	777	31.8	53.1
0.21-0.40	497	20.3	73.5
0.41-0.80	599	24.5	98.0
0.81 -1	50	2.0	100.0
Total	2445	100.0	

## REFERENCES

- [1] Collins, J. Writing Multiple-Choice Questions for Continuing Medical Education In Radiology, Activities and Self-Assessment Modules1 The Journal of continuing Medical Education, March (2006) 26, 543-551. PMID: 16549616
- [2] Tarrant M, Knierim A, Hayes SK, Ware J: The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. Nurse Educ Today 2006, 26(8):662-671. PMID: 17014932.
- [3] Tarrant M, Ware J: Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008, 42(2):198-206. PMID: 18230093
- [4] Tarrant, M, Ware, J M Mohammed, A (2009) An assessment of functioning and non-functioning distracters in multiple-choice questions: a descriptive analysis, BMC Medical Education 2009, 9:40-47 PMID: 1958068
- [5] Haladyna TM and Downing SM, Validity of Taxonomy of Multiple Choice Item writing Rules, Applied Measurement in Education, Volume 2, Issue 1, 1989 DOI: 10.1207/s15324818ame0201\_4
- [6] Shifflett, B., Phibbs, K., & Sage, M.. Attitudes toward collegiate level classroom testing. Educational Research Quarterly, (1997)21(1), 15-26.
- [7] Farley JK: The multiple-choice test: writing the questions. Nurse Educ 1989, 14(6):10-12 PMID: 15327679
- [8] McCoubrie P: Improving the fairness of multiple-choice questions: a literature review. Med Teach 2004, 26(8):709-712. PMID: 15763874
- [9] Vyas R, Supe A. Multiple choice questions: a literature review on the optimal number of options. Natl Med J India. 2008 May-Jun;21(3):130-3. PMID: 19004145
- [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 Ahmadabad, Gujarat. Indian J Community Med. 2014 Jan;39(1):17-20 PMID: 24696535
- [11] Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T. Item analysis of published MCQs. Indian Pediatr 1998;35:1103-5. PMID: 10216545
- [12] Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc 2012;62:142-7. PMID: 22755376
- [13] Rogers WT, Harley D: An empirical comparison of three- and four-choice items and tests: susceptibility to testwiseness and internal consistency reliability. Educ Psychol Meas 1999, 59(2):234-247, doi: 10.1177/00131649921969820
- [14] Swanson DB, Holtzman KZ, Clauser BE, Sawhill AJ: Psychometric characteristics and response times for one-best-answer questions in relation to number and source of options. Acad Med 2005, 80(10 Suppl):S93-96. PMID: 16199468
- [15] Zoanetti N, Beaves M, Griffin PI and Wallace EM: Fixed or mixed: a comparison of three, four and mixed-option multiple-choice tests in a Fetal Surveillance Education Program, BMC Medical Education 2013, 13:35 doi:10.1186/1472-6920-13-35
- [16] Rodriguez MC. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. Educ Measure 2005:3-13. [onlinelibrary.wiley.com/doi/10.1111](http://onlinelibrary.wiley.com/doi/10.1111)
- [17] Maha Al-Yahya, Ontology-Based Multiple Choice Question Generation, The Scientific World Journal, Volume 2014 (2014), pages -9 <http://dx.doi.org/10.1155/2014/274949>
- [18] Linn, R. L. & Gronlund, N. E. (1995). Measurement and assessment in teaching, Seventh edition . New York: Macmillan

<http://www.ejournalofscience.org>

- [19] Cronbach, L. J. My current thoughts on Coefficient Alpha and successor procedures. *Educational and Psychological Measurement*, (2004): 64, 391-418. <http://dx.doi.org.proxy2.library.uiuc.edu/10.1177/0013164404266386>

**Table 4:** Frequency distribution table of classified Cronbach's alpha

Cronbach's alpha	Frequency	Valid Percent	Cumulative Percent
under 20	4	10.0	10.0
0.20 -0.40	1	2.5	12.5
0.41-0.60	3	7.5	20.0

0.61-0.80	6	15.0	35.0
0.81 - 1	26	65.0	100.0
Total	40	100.0	-

**Table 5:** Frequency distribution of distracters according to selection

selected distracters	frequency	percent	cumulative
all	271	12	12
1-non-selected	778	34.6	46.6
2- non-selected	854	38.1	74.7
3 non-selected	342	15.3	100
total	2445	100	-