Biosciences, Biotechnology Research Asia Vol. **8**(1), 125-129 (2011)

Assessing the Psychometric Indices of Written Multiple Choice Educational Examinations of College of Paramedics of Kashan University of Medical Sciences in Education Year 2008-09

MANSOUR SAYYAH*, ZARICHEHR VAKILI[1] and MANZARDOKHT BIGDELI

Affiliated Faculty Membe of Educational Development Center (EDC)-
College of Medicine, Kashan University of Medical Sciences, Kashan (Iran).
*Corresponding Author

(Received: May 26, 2011; Accepted: June 28, 2011)

ABSTRACT

History and Objective

Evaluation of students' academic achievement is one of the most important affair in any educational setting. The majority of educational institutions use multiple choice exams to evaluate their students. The question items making up these exams need to possess certain psychometric properties if an exam is to be regarded as reliable instruments. Such property is amenable to examination by item analysis. The purpose of this research was to examine the multiple choice exams used in paramedic college of Kashan University of Medical Science in year 2008-9.

Material and methods

In this cross sectional descriptive study, the responses of students to the multiple choice exams as well as the correct responses to each item provided by the instructor in education year 2008-9 was used. All the multiple choices exams conducted in this period were used by codes so that the identity of the instructor as well as the student was kept confidential. The difficulty index, discrimination index, and Cronbach alpha was calculated for every exam and then means value for each index was calculated by LERTAL 5.0 software purchased from Assessment Systems Corporation of the United States to perform item analysis.

Statistical Analysis

A total of 1481 multiple choice question items presented to the students at the college of paramedics for evaluating 31 different subjects was analyzed. The results of analysis showed that the mean value of difficulty items, discrimination index and Cronbach alpha were 0.55, 0.36 and 0.82, respectively.

Discussion and Conclusion

The average of difficulty index of multiple choice exams of college of paramedic was 0.55. This value is within the recommended value of 0.6 to 0.7 purposed by Granlond (1985) and Nelson (2001) range 0.3 to 0.7. The average of discrimination index 0.36 is also higher than the 0.21 regarded as acceptable value by Nelson. The average of Cronbach alpha was acceptable for most of the exams, however, some of the tests had value less than the recommended value and needs careful re-examination. In addition, some of the items in the test had negative discrimination index and need to be revised in future tests. Further research is needed to construct more reliable tests.

Key words: Item Analysis, Difficulty index, discrimination index, Cronbach alpha, multiple choice exams.

INTRODUCTION

Evaluation of students' academic achievement is one of the most important and challenging affair in any educational setting. The majority of educational institutions use multiple choice exams to evaluate their students. The question items making up these exams need to possess certain psychometric properties if an exam is to be regarded as reliable instruments. There are researches indicating that college-level faculty does not write exams well (Guthrie, 1992;

[1]Corresponding Author: Faculty Member of Kashan University of Medical Sciences, Kashan (Iran).

Lederhouse & Lower, 1974; McDougall, 1997), and the side effects of such poor exams are reflected in student's performance by focusing on memorization only (Crooks, 1988; Shifflett, Phibbs, & Sage, 1997).

Different type of exams is used in educational institution including multiple choice and essay. Multiple choice tests are presently the most common and preferred types of tests that are in use in many educational settings. 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 psychometric properties of every item used in a multiple choice test. Item difficulty, item discrimination, and internal consistency are three important concepts in developing a good multiple choice exam. While difficulty index refers to the difficulty of an item for the respondents to correctly 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 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 these indices (Gronlund, 1985, Nelson, 2001, Ebel & Frisbie, 1986, Mehrens and Lehmann, 1991, Osterhof, 1990, Linn & Gronlund, 1995 and Hopkins, Stanley & Hopkins, 1990 ; Burch, et, al, 2008). For instance, Gronlund (1985) suggest item difficulty within the range 0.60 to 0.70 as an acceptable index for multiple choice exams while Nelson offers the range 0.30 to 0.70 as the desirable item difficulty.

High quality multiple choice items are difficult to construct but are easily and reliably scored. Ebel (1986) states that the item difficulty less than 0.20 for an multiple choice exam indicate the item is a poor item and believes that this level should not be less than 0.40 whereas Mehrens (1991) and Osterhof (1990) set a less restricted criterion and suggest the 0.20 to 0.40 as a sufficient level for an item to be included in a multiple choice exams. The internal consistency criterion knows as the Cronbach alpha is another index that is used to judge a multiple choice test. In this regard, different level for different test purposes has been offered. Linn (1995) states that the value for the internal consistency should be between 0.60 to 0.85 while Hopkins (1990) suggests that this value should be 0.90 or higher. Burch (2008) claims that it is necessary to determine reliability of a test for issuing certificate of competency for medical practice. In addition to the criterion described, when designing multiple choice test items, the distracters offered to the test takers are also important (Nelson, 2001). Placing a distracter or distracters that are chosen by none of the take testers reduces the number of alternatives and increases the likelihood of guessing an item correctly. For instance, in a four choice multiple choice test having two distracters that are not chosen by the test takes actually give a 50 percent chance of guessing an item correctly.

Considering the importance of such criterion in designing multiple choice exams, this descriptive research was designed to determine the item difficulty, item discrimination, internal consistency and distracters used in final examinations of college of health of Kashan University of Medical Sciences in education year 2008-9.

MATERIAL AND METHODS

This descriptive research was conducted in collaboration with education development center of education undersecretary of the university. All the 31 multiple choice exams given by the instructor at the college of health in education year 2008-9 were used as the data for item analysis by Laboratory of Educational Research1 Test Analysis Package (LRTAP version 5.0). Every exam was item analyzed separately by LRTAP and then the results of analysis of these 31 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:pc for further analysis. The results of all analysis were reported in appropriate tables.

RESULTS

Overall, 1481 multiple choice item in 31 exams for different subjects given by different

instructors were analyzed. The frequency of item difficulties were categorized according to what Nelson (2001) and other literature recommend. The difficulties index categories were set to less than 0.30, 0.30 to 0.70 and above 0.70. The result of this analysis is presented in table 1.

Table 1 shows that 8 percent of exams had item difficulty less than 0.30, 39.6 percent had difficulty index within the recommended range, that is, 030 to 0.70 and 52.3 percent of the exams had items difficulty over 0.70.

Table 1: Frequency distribution of classified difficulty index

Range	Frequency	Percent	Cumulative
Less than 0.30	301	3.20	3.20
0.30 - 0.70	703	5.47	8.67
0.71- 1	447	2.32	100
Total	1481	100	-

The 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 result of this analysis is presented in Table 2

Table 2: Frequency distribution of classified discrimination index

Range	Frequency	Percent	Cumulative
Negative to zero	233	15.8	15.8
0.0 to 0.20	245	16.6	32.4
0.40 -0.21	309	21	53.4
0.41-0.80	574	38.9	92.3
1 - 0.81	120	7.7	100
total	1481	100	-

As it can be seen in table 2, discrimination index for the items with negative or zero were 15.8% , between 0 to 0.20 were 16.6% , between 0.21 to 0.40 were 20%, between 0.41 to 0.80 were 38.9% and above 0,81 to 1 were 7.7%, respectively.

The frequency of this index for the entire test was classified into 5 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 result of this classification is presented in Table 3.

Table 3: Frequency distribution of classified cronbach index

Range	Frequency	Percent	Cumulative
Less than 0.20	1	3.2	3.2
0.20 - 0.40	1	3.2	6.5
0.41 - 0.60	2	6.5	12.9
0.61 - 0.80	4	12.9	25.8
0.81 - 1	23	74.2	100
total	31	100	-

In table 3, it can be seen that 3.2% of the exams had internal consistency less than 0.20 and 74.2% showed consistency index over 0.81 or more.

Finally, the distracters analyze showed that 22.7 percent of items contained all distracters that were sufficiently distracting to be selected by some respondents, while 40% of the items had one, 26.5% had two, 6.5% had 3, 4.3% had 4 nonelected choices. The result of this analysis is presented in table 4.

Table 4: Frequency distribution of selected distracters

Selected Distracters	Frequency	Percent	Cumulative
All	336	22.7	22.7
1-non-selected	593	40	62.7
2- non-selected	391	26.5	89.2
3 non-selected	96	6.5	95.7
4 non-selected	65	4.3	100
total	1481	100	-

Educational evaluation should not be limited to using a test or set of subtests to evaluate students but rather it should carefully employ test or tests that meet some sort of psychometrics characteristics. Under such circumstances, tests may render results that fairly classify student, diagnose their achievement or are used as the criterion to pass or fail students. In this research, all the multiple choice exams used at the college of health were evaluated by performing item analysis of each one of them separately. Results of this research showed that the average of item difficulty for the test conducted at the Paramedics College was 0.55. This value is approximately close to what Gronlund (1985) recommends and is with the range 0.3 to 0.70 that Nelson (2001) suggests. However, 32.2 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 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. The other index evaluated was the discrimination index. In this research, the average of discrimination index was 0.21. This value is with the range Nelson (2001) has suggested. However, 15,8 percent of items used in the exams showed negative discrimination values or values close to zero. Such item are not discriminating the good students from the weak ones plus do not account for the true total test variance. These items need complete revisions. The value of internal consistency may change by eliminating test items with low coefficient[11].

Finally, the distracter analysis revealed that 22.7 % of the all the distracters were sufficiently attractive to be selected where as 40% had 1, 26.5% had two , 6.5% had three and 4.3 had 4 nonselected distracters. This condition needs careful reevaluation since it indicates that nearly 40 percent of the questions in the exams were three choice, 26.5 were two choice and 6.5 were one choice items.

In summary, the results of item analysis of multiple choice tests used in the Paramedics College indicated that a considerable test items passed the criterion recommended by expert in the field, however, some test items were not sufficiently constructed to fulfill the objectives of the test takers. Further research and reevaluation of the test items may lead to improvement in test constructions by the instructors at this college.

REFERENCES

1. Ebel, R.L. & Frisbie, D.A., Essentials of Educational Measurement (4th ed.). Sydney: Prentice-Hall of Australia (1986).
2. Gronlund, N.E., Measurement and evaluation in teaching (5th ed.). New York: Collier Macmillan Publishers (1985).
3. Nelsonæ L. R., Item Analysis for Tests and Surveys, Curtin University of Technology (2001).
4. Kaplan, R.M. & Sacuzzo, D.P., Psychological testing: principles, applications, and issues. Pacific Grove, California: Brooks/Cole (1993).
5. Mehrens, W.A. & Lehmann, I.J., Measurement and evaluation in education and psychology (4th ed.). London: Holt, Rinehart (1991).
6. Oosterhof, A.C., Classroom applications of educational measurement. Columbus, Ohio: Merrill (1990).
7. Linn, R.L. & Gronlund, N.E., Measurement and assessment in teaching (7th ed.). Englewood Cliffs, NJ: Prentice-Hall (1995).
8. Hopkins, K.D., Stanley, J.C., & Hopkins, B.R., Educational and psychological measurement and evaluation (7th ed.). Englewood Cliffs, NJ: Prentice-Hall (1990).
9. Burch, V, C,. Norman, G, R,. Schmidt, H. G. Van der Vleuten, C. P,M., Are Specialist Certification Examinations a Reliable Measure of Physician Competence? Advances in Health Sciences Education, **13**(4): 521-533 (2008).
10. Basic Item Analysis for Multiple-Choice Tests. 1995-10-00 Basic Item Analysis for Multiple-Choice Tests. ERIC/AE Digest.
11. Guthrie, D. S., "Faculty goals and methods of instruction: Approaches to classroom assessment." In Assessment and Curriculum Reform. New Directions for Higher Education No. 80, 69-80. San Francisco: Jossey-Bass (1992).
12. Lederhouse, J. E., & Lower, J. M., Testing college professor's tests. *College Student Journal*, **8**(1): 68-70 (1974).
13. McDougall, D., College faculty's use of objective tests: State-of-the-practice versus state-of-the-art. *Journal of Research and Development in Education*, **30**(3): 183-193 (1997).
14. Crooks, T. J., The impact of classroom evaluation practices on students. *Review of Educational Research,* **58**(4): 438-481 (1988).
15. Shifflett, B., Phibbs, K., & Sage, M., Attitudes toward collegiate level classroom testing. *Educational Research Quarterly*, **21**(1): 15-26 (1997).