

A Training Program Based on the Quality Standards of Achievement Tests to Improve the Competence of Teaching Staff in Constructing Tests at Middle East University in Jordan

Basel Khamis Salem Abu Foudeh

Faculty of Arts and Sciences, Middle East University, Jordan

Received January 13, 2020; Revised February 18, 2020; Accepted February 24, 2020

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The aim of this study is to construct a training program based on the quality standards of achievement tests and to verify their effectiveness in improving the competence of the teaching staff in constructing tests at Middle East University. To achieve the objectives of the study, the researcher constructed a training program and assessment tool, which included 58 items distributed on five dimensions. The study sample consisted of (70) faculty members, who were randomly selected from different faculties during the academic year 2018/2019. The results showed a statistically significant difference ($\alpha = 0.05$) in the evaluation of achievement tests in favor of post measures, and revealed the effectiveness of the training program. The results showed no statistically significant differences ($\alpha=0.05$) in the evaluation of achievement tests attributed to specialization. The study recommended the adoption of the training program to compensate the lack of educational qualifications in one aspect of the evaluation process, and to take advantage of the quality standard list in the evaluation of the achievement test.

Keywords Educational Evaluation, Test Items, Training Program, Achievement Tests, Middle East University in Jordan

1. Introduction

Achievement tests play an essential role in a student's life in all stages of education. Based on their results, the future of a student is determined and many decisions are used against him.

The tests are still the primary means in evaluation of the educational process, as an aid in learning about the students' achievements of the educational goals, comparing the students with each other, diagnosing the strengths and weaknesses they have, or the selection of individuals for different professions.

However, most students hate tests, and avoid trying to appear for the experience in any way.

The reason may be that tests, like any other means of evaluation, may be used to achieve non-evaluation goals, such as punishment.

The other reason is the constant association with scores, and the consequent potential negative results such as failure (Odeh, 2010, p. 71; Abu Jarad, 2011, p. 90).

Another criticism of the tests is that they do not encourage improvement, but rather memorization (Murad and Soliman, 2005).

There are some negative aspects of some forms of tests used in the evaluation process, including: confining the role of education in training and exercising students to perform certain behaviors only, focusing on results rather than processes, and relying more on passive learning (Nixon, 1990).

Focusing on the scope and characteristics of the content rather than on the statistical characteristics of the test items requires reviewing prevailing practices in the design and construction of achievement tests (Allam, 2011).

The process of preparing achievement tests is carried out in a series of steps: defining the purpose of the test, analyzing the content, writing the teaching objectives, preparing the table of specifications, formulating the items, making the test, giving and correcting it, and finally analyzing and interpreting the results (Odeh, 2010, p. 146-147).

It is important to construct the table of specifications when designing the achievement test as a framework for obtaining a representative sample of the learning outcomes (Allam, 2011; Al-Noor, 2014; Odeh, 2010).

Procedures for examining and verifying achievement tests' information are considered vital if there is a desire for achieving justice for students in an educational institution (Lambert & Lines, 2000, p.4).

Many studies have been conducted in the construction of achievement tests and analyzing, such as: (Sinan, 2006; Said 2006; Saif, 2008; Algdah, Miqdadi & Abu Asfer, 2008; Galton, 2009; Ngozi, Agu & Aloysius, 2013; Al Nor, 2014; Al Miloz & Sherbini, 2015; Al Kaltham, 2015; Al-Tayeb, 2016; Al Oun, 2016; Al-Zaraa, 2016; Assaad and Wasila, 2017; Adaika, 2018; Abdul Wahab, 2019).

1.1. Study Problem and Questions

The researcher provided an evaluation study which included a sample of achievement tests in various facilities at the university, in order to evaluate and identify the degree of their compatibility with the quality of the achievement tests' construction standards.

It was noted that many faculty members lack the skills which are necessary for the preparation of achievement tests.

The ability to construct tests can be improved through various training programs. This study came to verify the effectiveness of a training program based on the quality standards of achievement tests to improve the competency of faculty members in constructing tests. So, the questions of the study are:

1. How effective a training program based on the quality standards of achievement tests can improve the competency of faculty teaching members in constructing tests at MEU?
2. Are there any statistically significant differences ($\alpha = 0.01$) in the evaluation of the quality of achievement tests prepared by faculty teaching members who have completed the training program due to their specializations (sciences, and humanities)?

1.2. Study Importance

- The present study acquires its theoretical importance from the importance of training of the faculty teaching staffs at the university in order to maintain an advanced level in constructing achievement tests. In addition, it makes sure that the faculty teaching members can state test items with acceptable psychometric properties.
- The study provides information and data for faculty teaching members at the university to introduce them the quality standards of achievement tests, training them to use these standards, and connect

their performances to contribute to the improvement and development of their performance.

- The present study provides a training program that contributes to the improvement of the competencies of faculty teaching members in constructing achievement tests.
- The study provides a tool that includes the most important criteria in constructing achievement tests, so that they can be used in evaluating achievement tests at the university.

1.3. Study Objectives

1. Construct a training program for faculty teaching members at the university according to the quality standards of the achievement test.
2. Measuring the effectiveness of the training program in improving the competencies of faculty teaching members in constructing the achievement test.

2. Methodology

The researcher used the quasi-experimental approach.

2.1. Study Population

The study population consisted of all (176) faculty teaching members at MEU for the academic year 2018/2019 according to the statistics of the Human Resources Department at the university.

2.2. The Study Sample

The sample of the study consisted of (70) faculty teaching members who were randomly selected and represented (39.2%) of the total society, of which (38) members in the scientific faculty, and (32) members in the humanities faculty.

2.3. Study Tool

A training program was developed based on the quality standards of the achievement tests (Odeh, 2010; Rawashdeh et al., 2000; Coop, 2006).

The training program consisted of (5) training sessions, (2) hours long, two sessions per week. The trainees were divided into groups, each group consisted of (4-6) members. The topics of the sessions were as follow:

- Session (1): Clarification of the overall objective of the program, definition of terms and concepts related to achievement tests, clarification of the role of achievement tests in evaluation, identifying the types of achievement tests and testing purposes, measurement of the learning outcomes, and producing the test sheet.

- Session (2): Steps to construct achievement tests, how to analyze the content, and how to construct a test's specifications table.
- Session (3): Writing multiple choice questions. The advantages and disadvantages, guidelines for their formulation, common mistakes while writing multiple-choice questions and how to correct these errors, displaying realistic and academic examples.
- Session (4): Right or wrong questions. The advantages and disadvantages, rules and guidelines for formulating them, some possible errors while writing these types of questions, displaying some activities and exercises.
- Session (5): Essay questions. The advantages and disadvantages, guidelines for formulating them, possible errors in writing these types of questions, and discussing some examples.

The validity of the training program was verified by consulting the specialized examiners. The researcher followed three stages to evaluate the training program:

- Stage (1): Preliminary evaluation of the program areas.
- Stage (2): Formative evaluation. It takes place during the application through tasks and assignments, and observing the views of faculty staffs and their interaction during the implementation of the program.
- Stage (3): Summative evaluation. This is done by writing an achievement test according to the test quality standards.

The researcher has prepared a tool to assess the quality of achievement tests which consisted of (58) standards that were derived from a review of theoretical literature and previous studies related to the international standards to construct achievement tests. The assessment tool consisted of (5) main dimensions as follows:

- Dimension (1): The standards for the output of the test sheet (14 standards).
- Dimension (2): The standards of constructing good achievement tests (12 standards).
- Dimension (3): The standards for formulating multiple-choice questions (11 standards).
- Dimension (4): The standards for formulating questions of right or wrong (10 standards).

- Dimension (5): The standards for formulating essay questions (11 standards).

Each of the above standards is given a score between (0-2).

- Score (2): If the given standard is completely performed.
- Score (1): If the given standard is performed to lower degree.
- Score (0): If the given standard is not performed at all.

The validity of the evaluation tool was verified by consulting (12) specialized examiners working in the field of educational measurement and evaluation.

The reliability of the tool was evaluated in its final form by applying it to (18) achievement tests.

The value of the coefficient of reliability of internal consistency; Cronbach alpha was (0.87). The values of reliability coefficients of the dimensions were between (0.83- 0.89).

3. Data Collection

Pre-test was performed by analyzing the study sample exams. Then the training program was implemented, and a questionnaire was distributed to assess the quality of the test sheet.

Post-test was performed by collecting the final exams prepared by the study sample and then analyzing them.

4. Results and Discussions

To answer the first question: How effective a training program based on the quality standards of achievement tests can improve the competency of faculty teaching members in constructing tests at MEU?

The results were prepared by the faculty members who completed the training program and compared with the results of the pre- performance.

The mean and standard deviations of the standards for each dimension of achievement tests were calculated. Table 1 shows the results of the analysis:

Table 1. Values of mean and standard deviations for pre and post-performance.

dimensions	NO.	Pre-performance		Post-performance	
		Mean	Std.	Mean	Std.
Standards for the test paper.	14	0.76	0.27	1.92	0.21
Criteria for building a good achievement test.	12	0.68	0.25	1.89	0.25
Criteria for formulating multiple choice questions.	11	0.81	0.16	1.89	0.26
Criteria for formulating questions of right or wrong.	10	0.47	0.21	1.88	0.22
Criteria for drafting essay questions.	11	0.92	0.12	1.85	0.25
Total	58	0.73	0.16	1.89	0.20

It is noted from the results of table 1 that there are apparent differences between the values of the means in the pre- and post-measures.

The total mean of the pre-performance is (0.73) with a standard deviation of (0.16), while the value of the mean of the post-performance is (1.89) and standard deviation (0.20).

To determine the significance differences between the values of the means, the researcher used a paired sample t-test, and the effect size was calculated using the following equation:

$$\text{Eta Square } (\eta^2) = t^2 / (t^2 + df)$$

Where:

η^2 :Eta Square. t^2 : square t-value.

df: degrees of freedom.

To determine the effect size (the ratio of the total variance in the dependent variable, which can be attributed to the independent variable). The following values were adopted (Kieiss, 1989, p. 448):

- The value of the Eta-Square is less than (0.20); the effect size is weak.
- The value of the Eta Square between (0.20-0.80); the effect size is medium.
- The value of the Eta-Square is greater than (0.80); the effect size is high.

The effectiveness of the training program was calculated using Modifies (Black Gain Ratio) as follows (Al-Jarallah, 2013, p. 87):

$$\text{Black Adjusted Gain Ratio} = \{(Y - X) \div (d - X)\} + \{(Y - X) \div (d)\}$$

Where:

Y: mean of post-performance scores. X: mean of pre-performance scores. d: Final score of the assessment.

The range is determined to measure effectiveness based

on the percentage set by Black for the effectiveness which is placed between (1-2).

Black suggests the boundary of this ratio is (1.2%), so that the effectiveness of the program can be considered acceptable. Table 2 shows the results of the analysis.

It is noted from the results of table 2 that there is a statistically significant difference ($\alpha=0.01$) attributed to the post-performance; that is, the sample performance improved after training.

All values of Eta-square (η^2) were large, indicating that the value of the effect of the experimental variable (training program) was large. The adjusted gain ranges between (1.33-1.63), and all values above (1.20%).

The adjusted total performance gain was (1.48), which is large. This indicates the effectiveness of the training program.

It can be said that the proposed training program improves the skills of faculty staffs in constructing achievement tests, as it meets the needs of faculty staffs in all aspects of constructing achievement tests.

The training program focused on the cognitive and practical aspects of the stages of constructing the achievement test. It offered diverse activities and teamwork, as well as the trainees' desire to be benefited to the fullest extent from the development of achievement tests in accordance with the quality standards.

The results are consistent with the studies (Al Tayeb, 2016; Al-Zaraa, 2016; Abdul Wahab, 2019).

To answer the second question: Are there any statistically significant differences ($\alpha = 0.01$) in the evaluation of the quality of achievement tests prepared by faculty teaching members who have completed the training program due to their specializations (sciences, and humanities)?

Independent sample t-test analysis was used. Table 3 shows the results of the analysis.

Table 2. The results of the t-test in the pre and post measures and the adjusted gain

Dimensions	Means		The difference between Means	t-value	df	Sig.	Eta Square (η^2)	Effect size	Gain Ratio
	Pre	Post							
Standards for the test paper.	0.76	1.92	1.16	28.81	69	**0.0	0.92	High	1.52
Criteria for constructing a good achievement test.	0.68	1.89	1.21	30.07	69	**0.0	0.93	High	1.52
Criteria for formulating multiple choice questions.	0.81	1.89	1.08	29.37	69	**0.0	0.93	High	1.45
Criteria for formulating questions of right or wrong.	0.47	1.88	1.41	37.01	69	**0.0	0.95	High	1.63
Criteria for drafting essay questions.	0.92	1.85	0.93	26.16	69	**0.0	0.91	High	1.33
Total	0.73	1.89	1.16	36.82	69	**0.0	0.95	High	1.48

** Statistically significant at the level ($\alpha = 0.01$).

Table 3. The results of the t-test

Dimensions	specializations	NO.	Mean	Std. Deviation	Df	t	Sig.
Standards for the test paper.	Sciences	38	1.90	0.24	68	0.836	0.406
	Humanities	32	1.95	0.17			
Criteria for building a good achievement test.	Sciences	38	1.90	0.25	68	0.262	0.794
	Humanities	32	1.89	0.26			
Criteria for formulating multiple choice questions.	Sciences	38	1.90	0.26	68	0.173	0.863
	Humanities	32	1.89	0.26			
Criteria for formulating questions of right or wrong.	Sciences	38	1.90	0.20	68	0.843	0.402
	Humanities	32	1.86	0.23			
Criteria for drafting essay questions.	Sciences	38	1.87	0.25	68	0.579	0.565
	Humanities	32	1.84	0.26			
Total	Sciences	38	1.90	0.21	68	0.191	0.849
	Humanities	32	1.89	0.19			

** Statistically significant at the level ($\alpha = 0.01$).

It is noted from the results of table 3 that there are no statistically significant differences ($\alpha=0.01$) in assessing the quality of achievement tests attributed to specialization, where all values of "t" were not statistically significant ($\alpha=0.01$).

This result can be explained by the nature of cognitive competencies related to constructing achievement tests, which are acquired through learning, training and experience.

The results are consistent with the study of (Saif, 2008; Algdah, Miqdadi & Abu Asfer, 2008; Al Miloz & Sherbini, 2015). Studies have shown that there are no statistically significant differences attributed to specialization.

5. Recommendations

1. The study suggested adding preparation programs for faculty members before and during service to the list of quality standards of achievement tests because of its positive impact in improving the quality of their professional performance.
2. Adopting the training program to compensate the lack of educational qualification in one aspect of the evaluation process.
3. Utilizing the list of quality standards of achievement test in evaluating tests prepared at the university.
4. Making studies to develop training programs in all aspects of the measurement and evaluation at the university.

Acknowledgments

The author is grateful to Middle East University, Amman, Jordan for the financial support granted to cover the publication fee of this research paper.

REFERENCES

- [1] Adaika, Dunya. Level of Knowledge of Secondary Education Teachers for the Qualifications of Preparing Good Achievement Tests in the Light of Their New Roles: An Exploratory Study at the Central Valley High School, *Journal of the Researcher in the Humanities and Social Sciences*, Vol. 3, No. 33, 1055-1072, 2018.
- [2] Abdul Wahab, M. Training program in writing multiple choice items and its effect on the attitude towards achievement test construction, *Educational Journal in Kuwait*, Vol. 33, No. 130, 197-240, 2019.
- [3] Allam, Salahuddin. Educational and psychological measurement and evaluation, Cairo, Dar Al-Fikr Al-Arabi, 2011.
- [4] Assaad, Sabah and Wasila, Ben Amer. Evaluating the competency of constructing the Achievement Tests for University Teaching Teachers in Accordance with the Good Test Standards: An Analytical Study of the Achievement Tests for the First and Second Semester of the University Years 2012-2013, *Journal of Humanities and Social Sciences*, Vol. 28, No. 81-90, 2017.
- [5] Abu Jarad, H. The extent of teachers' compliance with achievement tests analysis and its relation with their attitudes toward it, *Journal of Azhar University-Gaza*, Vol. 13, No. 2, 89-106, 2011.
- [6] Al-Jarallah, Nada. The effectiveness of teaching biology using the realistic model in the achievement and developing critical thinking skills of first year secondary students, Unpublished Master Thesis, Curriculum and Methods of Teaching Science, King Khalid University, College of Education, 2013.
- [7] Al-Miloz, H. and Sherbini, Gh. Quality standards of e-examinations as perceived by faculty members and students in the faculties of girls at King Khalid University, *International Interdisciplinary Journal of Education*, Vol. 4,

No. 4, 25-42, 2015.

- [8] Al-Zaraa, Layla. The effectiveness of a training program in improving the competencies of constructing achievement tests for faculty members at King Faisal University, *Journal of Educational and Psychological Sciences*, Vol. 17, No. 2, 277-304, 2016.
- [9] Al-Gdah, M., Miqdadi, H., and Abu Asfer, R. The extent of applying the standards of designing a good test from the viewpoint of faculty members and students at the Teachers College in Abha, *Yearbook of Teachers College in Abha*, Vol. 13, 274- 302, 2008.
- [10] Al-Kaltham, Hamad. Evaluating some Tests of Islamic Education Umm Al-Qura and Al - Dammam Universities in the light of the test quality standard, *Journal of Arab studies in education and Psychology*, Vol. 66, 209-236, 2015.
- [11] Al-Tayeb, Badawi. The effectiveness of a training program in developing the skills of formulation achievement tests for reading teachers and their impact on their students in the framework of the preparation for international study (PIRLS), *Journal of Reading and Knowledge*, Vol. 172, 21-75, 2016.
- [12] Al-Oun, Ismail. Efficiency of teachers of physical education in building achievement tests in the schools of education in Northeast Badia of Jordan according to the standards of a good test, *Association of Arab Universities journal for education and psychology*, Vol. 14, No. 3, 176-213, 2016.
- [13] Al-Nor, Ahmad. The evaluation of criteria of the good achievement test applied by the Arabic language staff at Sudan University of science and technology according to students' point of view and other variables, *International Interdisciplinary Journal of Education*, Vol. 3, No. 5, 1-21, 2014.
- [14] Coop, I. Increasing Teacher Effectiveness: Research and Practice, *International Journal of Science Education*, Vol. 13, No. 2, 27- 47, 2006.
- [15] Galton, Azhar. Evaluating the final exams for graduate studies at the Faculty of Education Sana'a in the role of good testing standards, *Scientific Journal, Faculty of Education*, Vol. 8, 187-200, 2009.
- [16] Lambert, D., and Lines, D. *Understanding Assessment*, London: Routledge Falmer, 2000. Murad, S. and Soliman, A. *Tests and Measurements in Psychological and Educational Sciences: Steps Preparation and Characteristics* (2nd ed.), Cairo, Dar Modern Book, 2005.
- [17] Negozi, N., Agu, c., and Aloysius, C. Measuring teachers' competencies in constructing classroom - based tests in Nigerian secondary schools, *Need for a test construction skill inventory*, *Academic Journals*, Vol. 8, No. 8, 431-439, 2013.
- [18] Nixon, N. Assessment issues in relation to experience based learning on placements within courses, In C. Bell and D. Harris (eds.), *Assessment and Evaluation World Yearbook of Education*, London, Kogan Page, 1990.
- [19] Odeh, Ahmad. *Measurement and Evaluation in the Teaching Process* (4th ed.), Irbid, Dara Al Amal, 2010.
- [20] Rawashdeh, Ibrahim et al. *Teacher guide in constructing achievement tests*, Ministry of Education, Amman, Jorda, 2000.
- [21] Said, Al-Saeed. Evaluation of the questions of open education exams, Faculty of Agriculture. Research presented at the Eighteenth Scientific Conference, Concepts of Education and constructing the Arab Man, Egyptian Society for Scientific Education, July, Ain Shams University, 2006.
- [22] Saif, Muhammad. The level of knowledge of the principles of the supervisors' constructing achievement tests in the province of Taiz Governorate. Unpublished Master Thesis, College of Educational Sciences, University of Jordan, Amman, 2008.
- [23] Sinan, Enas. The degree of knowledge and practice of mathematics teachers in the intermediate stage of the skills of constructing achievement tests in the city of Makkah, Unpublished Master Thesis, College of Education, Umm Al-Qura University, 2006.