Multicultural Education
ISSN: 10683844
Volume 6, Issue 1, 2020

The Impact Of The Test Length On The Accuracy Of Estimates Of The Ability Indicators For Individuals and The Information Function Of The Test According To The Item Response Theory

Ahmad Abdel Hafed Alqraleh

Article Info

Article History

Received:
June 12, 2020

Accepted:
August 15, 2020

Keywords
Test Length,Item Response Theory, Information Function, Ability

DOI:
10.5281/zenodo.3992202

Abstract

This study aimed at verifying the impact of the test length on the accuracy ofestimatesof the ability indicators for individuals and the information function of the test according to the item response theory, to achieve the objective of the study, a multiple-choice test in chemistry was constructed for students of the 10th grade in Jordan. The test is made up of three models of the same content, but vary in thenumber of items, the extendedtest, the intermediate test, and the short test. The(Bilog-Mg) program was used to analyze the 915 students' responses for all three test models according to the two-parameterlogisticmodel. The study results revealedno statistically significant differences between the means of the standard errorfor estimating individual ability indicators attributed to the length of the test. The results of the study showed that the information function of the test differed with the variation of the test models, as well as the absence of statistically significant differences attributed to the values of the empirical validity coefficients.

1. Introduction

Since the foundation of the psychological movement, psychologists have been interested in achieving the highest degree of objectivity in tests while using tests and psychological scales. The field has witnessed increasing developments designing, Constructing, and analyzing test items to achieve this objective (Abu Hatab, 1992). With criticism of classical measurement theory (CCT), contemporary measurement scientists have made innovative research efforts since the 1970s to develop a modern sociometric theory that can overcome many of the traditional measurement problems called the Item Response Theory (IRT) or Latent Traits Theory (LTT). This theory was evolved, and multiple models of significant development emanated since that time until the present time, these developments were mainly based on the advanced technology and computer software, which became available to researchers in several areas (Allam, 2005). In the item response theory, psychological and educational tests generally assume that certain traits or characteristics are shared among all individuals, but vary in size. Although these traits are unobservable, it can be inferred from the observed behavior of an individual whose responses to the test items justifies its designation as latent traits (Allam, 1987).

`The item response theory consists of a set of mathematical models showing how examinees respond from different levels of ability to test items. These models are also used to estimate the item's and individuals' characteristics in these traits through several indicators. Using the estimated values of these indicators can explain the performance of each individual in the test, Using the estimated values of these indicators, we can interpret the performance of each individual in the test, and because these features are difficult to observe and measure directly, they are inferred by using these estimated values, they are called latent traits (Salah al-Din Allam, 2005, p. 47; Crocker & 1986, p450).

Psychologists have been interested in achieving the highest degree of objectivity when using psychological tests and measurements. To achieve this goal, this field has witnessed increasing developments in designing, constructing, and analyzing test items (Abu Hatab, 1992). The testing is one of the essential tools that provide us with the data used to make many important decisions related to the individual and society. The test is a systematic procedure for measuring a sample of the behavior of individuals. The use of tests has been widespread in many fields. Such as choosing a person for a job from among a group of applicants, or for classification purposes such as determining the students' path in proportion to their abilities and skills, as well as evaluating the achievement of students through the grades obtained in classroom tests, etc. (Allen & Yen, 1979). The items of the multiple-choice test are considered the most flexible among tests. It can be used effectively in the appropriate coverage of the educational material, and the ease and objectivity of the correction because it requires the identification of the correct answer, unlike the essay questions that require recalling answer, they were known for their

accuracy, reliability, their correctness and is not affected by the subjective factors of the corrector (Samarah et al., 1989).

The multiple-choice test items are considered the most flexible among all tests; this type of tests can be used effectively in the appropriate coverage of the educational material.Moreover, it is the ease and objectivity of the correction because it requires the identification of the correct answer, as correctors never dispute the answers, unlike the essay questions that require recalling answers. They were known for their accuracy, reliability, validity, and is not affected by the subjective factors of the corrector (Samarah et al., 1989).

Because of the widespread of these tests and their importance in several areas, one of the essential aspects of the tests is the length of the test, as some consider the best test is the extended test, others prefer the medium length test. In contrast, some consider the short test is the best, hence the importance of the study to examine the impact of the length of the test on the accuracy of the individual ability estimates and the information function of the test according to the binary parameter model, one of the models of the item response theory.Despite the interest of psychometric researchers and educational scientists in several aspects of the tests in general, but the subject of this study did not receive excessive attention by the scientists of psychological and educational measurement, there is a paucity in the studies that examined the subject of this study. The study of (Al-Omari, 1985) is one of the studies related to the subject of this study, which aimed to know the impact of the length of the test time on the performance of the examinees, and its psychometric characteristics such as,validity and reliability among examinees related to the general concern and risk levels when taking the test using multiple-choice items. The test consisted of (25) question; each question has (4) alternative; only one is correct. The results of the statistical analysis of the binary comparisons of the differences between the reliability coefficients of the test showed a differences between the reliability coefficients of the test attributed to the length of the test. Among the studies related to the subject of the study is the study of (AL-Diyabat, 2007), which aimed to identify the effect of the length of the test on the reliability of the test. The study results showed the differences between the test attributes' reliability estimates and the change in the length of the test. The results indicate that the reliability estimates increase with the length of the test, as the total test had the highest reliability coefficient estimate (0.9585). The least length had the lowest reliability coefficient estimate (0.832).

It is clear from the previous studies that the effect of the test's length on the accuracy of the individuals' ability estimates and the test's information function of the analysisshows that there is a difference in the results of these studies according to the Modern Theory of measurement.

1.2 Problem Statement

The tests are one of the essential tools that provide us with data based on making many important decisions concerning the individual and society. The use of the criteria has been widespread in many fields. These tests are designed for a variety of purposes such as selecting a person for a job from a group of applicants, or for classification purposes such as determining the students' path in proportion to their abilities and skills, as well as in evaluating the achievement of students through grades they obtained in the classroom tests,etc. Because of the importance of these tests, this study was designed to examine a specific aspect of the tests, namely, to identify the impact of the test's length on the accuracy of the individual ability estimates and the information function of the test according to the modern theory of measurement

Therefore, the current study used the Item Response Theory to analyze the response results on the three test models, which provided many solutions to the problems related to the construction and development of the tests. In modern theory, the estimates of the individual's abilities are free from item characteristics.The forecastfor item characteristics are free from the ability of the individuals sample; specifically, this study seeks to answer the following questions:

Question 1: Does the accuracy of the estimated ability rating of the estimated individuals using the binary parameter model differ depending on the length of the test?

Question 2: Do the values of the information function of the test differ depending on the length of the test?

1.3 The importance of the study

The importance of the study is that it aims to identify the impact of the length of the test on the accuracy of the estimates of the individual ability indicators and the information function of the test. The particular response theory was used in analyzing the data derived from the application of the three test models, which vary in the item number. Where the indicators of the ability of individuals were estimated independently of the individual's skills, in addition to the results that will be provided by this study from

the teachers and official bodies existed and interested in this area, as well as in the construction and development of tests from the competent authorities in the field of education, This study is considered as a modest scientific addition to the Arab library, through the results of this study there can be further research using the accumulation of knowledge in this area, through further education and similar analysis on a broader scale, and in other subjects, thus forming a link between past and future studies.

1.4 Study Limits

There are some determinants to be observed when circulating results outside the study community, represented in:

1-The results of this study are limited to the study sample represented by students of the tenth grade in Jordan.

2- The study tool is limited to the achievement test of the primary grade 10 in chemistry in Jordan.

3- The results of this study are determined by the two-parameter logarithmic model (2PLM), one of the item response theory models.

1.5 Definitionof Study Terms

1.5.1Two-Parameter Logistic Model

One model of individual response theory, based on two indicators of the individual: the index of difficulty, and the index of discrimination.

1.5.2Test Information function

It is a mathematical function that expresses the sum of the functions of the test items.

1.5.3Estimation of an individual's capacity parameter (Person's parameter)

It means estimating the ability parameter for each examinee by a value derived from the application of a mathematical function in one of the item responses models in response to the test subjects.

1.5.4 The length of the test

The number of items that comprise the test. In this study, three tests were formed (long test, intermediate test, and short test).

2. Previous Studies

Alqudah& Al-Shraifin (2020) study aimed at identifying the effect of the length of the questionnaire on the Accuracy of Ability Estimation of the parameters of examinees and the parameters of Items and scale in the light of IRT. The results showed statistically significant differences at the level of ($\alpha = 0.05$) between the standard error average in the estimation of the abilities of individual related to the length of the questionnaire; the results showed differences between the standard error averages in favor Length of questionnaire

Al-Qaisi (2016) study aimed at investigating the effect of the sample size and length of test on the accuracy estimation of the item-parameter by using the Non- Parametric item- response theory. The Findings showed statistically significant differences at ($\alpha = 0.05$) in the means of Bias in the estimation of the ability parameter θ attributed to the (Sample size and test length).

Allen (2016) stated that many factors influence the response rate of a survey or questionnaire. The BYU alumni questionnaire was initially a lengthy survey with over 200 questions. After a short version of the questionnaire was created and administered, response rates appear to have increased substantially. Male respondents appear particularly more inclined to respond to the shortened version compared to the long version.

Alhawari (2015) study investigated the Effect of Test Length and Ability distribution form on The Estimation of a person's Ability, item difficulty, and the information function of test and its items, According to Rasch Model in Item Response Theory (IRT). The results showed that there were statistically significant differences at ($\alpha=0.05$) among the standard error means of item in the estimation of difficulty parameters, such estimations in a person's ability were more accurate in the positive and negative skewed. The test consisted from 30 items.

3. Methodology

3.1 Population of the Study

The study population consists of all the regular tenth-grade students in the Jordanian public schools. Karak governorate was chosen to be the target population of the study, which includes four directorates of education: the Directorate of Education of the Karak Region, the Directorate of Education of the Qaser Region, the Directorate of Education of AL-Mazar Region, the Directorate of Education of the Southern Ghour Region. The number of the target community was (16640) male and female students, according to statistics available at the Ministry of Education during the first semester of the academic year 2014/2015.

3.2 Study Sample

The researcher selected the study sample by a random cluster sample method at the tenth-grade level students in Karak governorate.The sample consisted of 945 students. The sample of each of the three models of the test was 315 students. The three test models were randomly distributed on the study sample to ensure the chances of parity between them.

3.3 StudyTool

To achieve the study's objectives, the researcher prepared an achievement test in chemistry for students of the tenth grade in Jordan. The experiment consisted of (60) items of multiple-choice type, with four alternatives for each item. The researcher in the writing of articles has adopted the general principles used by (Gronlund & Linn, 1990) in the construction of tests of achievement. The following are the procedures followed by the researcher:

3.3.1 To determine the purpose of the test

It is a measure of the level of students' achievement in chemistry at the tenth-grade level in Jordan.

3.3.2 Analysis of the content of the study

The content of the chemistry subject was analyzed into (concepts, terminology, symbols, generalizations, skills, and applications) for each of the chemistry subject units of the basic10th grade in Jordan.

3.3.3 Building the specification table

In which the levels of educational outcomes were linked to the content of the subject of the test.

3.3.4 Formulation of the test item

(60) items wereformulated of the multiple-choice type, four alternatives for each item, one representing the correct answer.

3.3.5 Validity of the test content

The content of the test was presented in its preliminary form, as well as the content analysis, the educational objectives, and the specification table to several experienced and competent arbitrators. Based on the arbitrators' notes, some words such as ambiguous words and weak substitutes were amended, which were not suitable for the achievement of the expected objective.

3.3.6 The initial experiment of the test on the sample survey

After the formulation and modification of the test items according to the views of the arbitrators, the test was printed in its final form, consisting of (60) element, and then applied to a group of (82) male and female students other than the study sample, to get the preliminary analysis of the test elements and the detection of the items that need to be modified or deleted by identifying the level of difficulty and its discrimination ability as well as the calculation of the reliability coefficient of the test, and the capabilities of students represented by their marks on the test, besides collecting any notes about the test items. A separate paper has been prepared to answer the test to get one score for each correctly answered item. Thus, the total mark of the student is the sum of the correct answers.

Difficulty and discrimination coefficients were calculated for each item using the statistical program (SPSS).

Table 1: shows the difficulty and discrimination coefficients of all test items for the exploratory sample

item	Diff	Discr	item	Diff	Discr	item	Diff	Discr	item	Diff	Discr

1	0.75	0.84	16	0.51	0.52	31	0.41	0.52	46	0.75	**0.44**
2	0.33	0.77	17	0.61	0.61	32	0.38	0.39	47	0.33	**0.33**
3	0.41	0.66	18	0.75	0.44	33	0.48	0.60	48	0.41	**0.50**
4	0.54	0.30	19	0.33	0.33	34	0.39	0.55	49	0.54	**0.60**
5	0.43	0.53	20	0.41	0.50	35	0.33	0.51	50	0.43	**0.61**
6	0.55	0.34	21	0.54	0.60	36	0.48	0.62	51	0.55	**0.42**
7	0.57	0.54	22	0.42	0.64	37	0.41	0.49	52	0.57	**0.40**
8	0.48	0.79	23	0.51	0.41	38	0.38	0.39	53	0.48	**0.62**
9	0.41	0.66	24	0.40	0.49	39	0.48	0.60	54	0.41	**0.52**
10	0.33	0.58	25	0.56	0.40	40	0.62	0.36	55	0.33	**0.33**
11	0.48	0.70	26	0.42	0.53	41	0.51	0.52	56	0.48	**0.60**
12	0.52	0.62	27	0.51	0.52	42	0.61	0.61	57	0.52	**0.49**
13	0.44	0.72	28	0.86	0.44	43	0.65	0.67	58	0.44	**0.60**
14	0.35	0.55	29	0.53	0.53	44	0.75	0.45	59	0.35	**0.31**
15	0.39	0.82	30	0.47	0.43	45	0.77	0.55	60	0.39	**0.47**

Table 1 shows that the difficulty coefficients' values for the initial view of the test ranged from 0.30 to 0.86. These coefficients are considered to be reasonable and appropriate. The table also shows that the average value of the test items difficulty coefficients was (0.50). As for the values of the discrimination coefficients of the test items, they range from (0.65 -0.31), and the mean value of these indicators was (0.52), which are good values and also acceptable for the study purposes.

As for the reliability of the test, the researcher evaluated the reliability value of the chemistry test in its original form, which consists of (60) item, by calculating the coefficient of the internal consistency value estimated using the equation Koder-Richardson-20 (KR-20) (0.92) on the experimental study sample which consists of (82) male and female students, the reliability value calculated using this method was(0.92),this value is considered high and a good indicator of the reliability of the test in measuring the target attribute.

After the previous steps, which confirmed the availability of the test's psychometric characteristics of the test, three models were formed as follows:

A) Extended test: A multi-choice test consisting of (60) items, and the examinee must choose the correct answer among them.

B) Medium test: A multiple-choice test consisting of (45) items, after the deletion of (15) items, from the long model, and the examinee must choose the correct answer among them.

C) Short test: A multi-choice test consisting of 30 items after deleting of (15) individuals from the medium model. The examinee should choose the correct answer among them.

3. 4 Basic Study Procedures

After the study sample was selected, the researcher visited the selected schools for applying the study tool. The three test samples were randomly distributed to the study sample. After the application was completed, the correction was done manually by the researcher using the sample response paper prepared for each sample of the three test models, One grade was given to each item of the achievement test when the examinee answers correctly, and zero(0) grade in the case of wrong answers, so that the number of grades or marks achieved for each student for each sample of the test is equal to the number of items of that model which the student answered it correctly, The answer papers of the sample of the study wasthen inspected, the number of unanswered answer papers was excluded, as well as those that did not bear the seriousness of the answer. The responses of (930) students were kept, each model containing (310) respondents, then collected data were entered into the computer to analyze the data through Bilog-

Mg3 and SPSS software to examine the assumptions required by the Binary Logistics Model, the item response theory hypothesis, according to the following steps:

4. Main Results

4.1 Unidimensional

A unidimensional assumption of the current study was verified by conducting a factor analysis of the individual responses on each of the three test models. The principal component analysis method employed using SPSS. The first and second latent root values and the ratio between them, as well as the percentage of variance explained by each factor, shown in Table (2).

Table (2): latent root values and the variance ratio explained in the different distributions of the test item

Test Model	Latent Root	Factor	Variance
Long	9.23	1	20.62
	3.87	2	8.08
Medium	11.33	1	22.53
	3.99	2	9.31
Short	12.25	1	25.03
	3.91	2	4.51

Table (2) shows that the ratio between the value of the first latent root and the second latent root of each of the three test models was greater than (2), it is an indicator to a uni-dimensionality of each of the three test models. Through these results, we find that the three test models have achieved the first assumption of the Item Response Theory.

4.2 Local Independence assumption

That the assumption of local independence is equivalent to a uni-dimensional assumption, which implies that the assumption of local independence of the test of the present study instrument has been achieved after a uni-dimensional assumption has been ascertained, as recommended by Hamilton and Swaminathan &Rogars, 1991.

4.3 Goodness of fit test

To examine this hypothesis, the researcher used the Bilog-Mg3 program, which is used to analyze data according to individual response theory. The items and individuals that did not match the bilingual logistic model for each of the three models were identified. Items and individuals that did not match the model were excluded according to the following steps:

4.3.1 delete the non-conforming individuals of the used binary parameter model by analyzing the data for each of the three test models and using a useful match statistic (k2 test at the level of $\alpha = 0.01$) so that any individual that does not fit to the binary model used is deleted if the probability value is less than ($\alpha = 0.05$),

The results of the matching of the first model showed that two individuals,two individuals, from the second model and five individuals from the third model of the test did not match the expectations of the bilingual logistic model, therefore their responses were deleted from each of the three test models, (915) male and female students responses were retained, each model contains (305) individuals.

4.3.2 delete the non-conforming items of the used binary parameter model by analyzing the data for each of the three test models and using a useful match statistic (k2 test at the level of $\alpha = 0.01$) to delete any non-matching binary model used if the probability value was less than ($\alpha = 0.05$). The results of the data analysis of the three test models showed that none of the items of the three models was deleted. After this step, all three components of the three test models were considered to be identical to the binary model.

4.4 Speed free

The assumption of speed is implicitly assumed in a uni-dimensional hypothesis. Therefore, the researcher provided sufficient time for the examinees to answer the study instrument represented in the chemistry test. The test time was determined on the basis of the initial experimental sample to achieve the assumption speed free.

4.5 Statistical Methods

Programs usedStatistical software statistical package (SPSS)(BILOG-MG) was used for data analysis, extraction, and calculation of:

- Calculating the internal consistency reliability coefficient3 value estimated in the Koder-Richardson 20 equation (KR-20).
- Assessment of the ability of examinees.
- Test information function for the three test images.
- Use the method of analysis of the uni-variance to answer the questions of the study.

4.6 Results and discussion

The present study aimed to determine the effect of the length of the test on the accuracy of the estimation of the ability indicators and the test'sinformation function according to the modern theory of measurement. The following are the results of the study.

Question 1: Do the estimates of ability indicators of individuals estimated using the binary model differ depending on the length of the test.

To answer this question, the free values of individual capacities and standard errors were estimated for each of the three test models according to their length. The mean ability indicators for the extended test was(0.0105), for the medium test was (0.0202) and for the short model (0.0286), the values of the standard errors means were acceptable for all test models,

This indicates the accuracy of the trait estimation for most individuals in each model. to detect the differences between the means of standard errors in estimating individuals'abilities for the three test models. Analysis of mono-variance as shown in the table (3).

Table (3): Results of the analysis of the mono-variance of standard error mean estimating the ability parameters of the three test's length differences.

Source of variance	Total squares	Freedom grades	Mean squares	F value	Statistical significance
between groups	0.001	2	0.000	0.007	0.822
Inside groups	21.030	6.4	0.031		
Total	21.030	6.6			

The results in Table (3) indicate that there are no statistically significant differences at $\alpha = 0.05$ level between the mean, standard errors in the estimation of individual's abilities attributed to the difference in the length of the test. Previous results can be explained by the fact that the test models applied to the examinees are the same, as the text of the items doesn't change, However only the length of the test that changed, and these changes did not affect the estimates of the capacity indicators of the individuals examined because all students studied the same content of the test material in various teaching methods with enrichment activities.

Question 2: Do the information function values of the test differ with the length of the test?

to answer the third question, the information function for each of the three test vocabulary items was obtained according to the binary model, the following are; the results:

First: The function of the item's information of the three test models

The binary model was used to obtain the highest and lowest value of the three test model profiles according to the time, Table (4) shows this.

Table (4) the values of items information for the three tests according to length, range, and arithmetic mean

Type of test	Highest information value	Lowest information value	Range	Mean of information
long	2.35	0.01	2.34	0.90
medium	3.22	0.07	3.1	0.91
short	5.09	0.19	4.90	0.95

Table (4) shows that the highest information value for large test items was (2.35), for intermediate information value (3.22), for low test value (5.09), the short test model was the highest value comparing to first and second models. The mean of the short test was the highest among all of the three tests (0.95), comparing with the first and the second tests (0.90) and (0.91), respectively.

Second: The information function of the three test types according to the difference in time.

A function information curve of the three test forms was plotted according to the length of the test, by identifying the value of the information provided by each form of a test at each point on the continuum of the capacity, through which the values of the information function and the corresponding coefficients of the reliability of the three test forms were obtained according to the test length as in table (5).

Table (5):The information function values and the corresponding index of the form's reliability three test forms according to the time

Form	Reliability Coefficient	Information function value
Short	0.82	23.25
Medium	0.85	26.40
Long	0.91	30.02

Table (5) shows that the highest reliability coefficient for the three test forms was for the third form (0.91) followed by the second form (0.85) and finally, the first form (0.82). To find out the significance of the differences in the reliability coefficients calculated from the item response theory, the proposed (M) test was used by Hackstin& Whalen (1976), which follows the distribution of the Kai square with freedom degree (1-number of coefficients), where the calculated (M) value was 1.754, which is less than the critical value of a quadratic square in the degrees of freedom (2), thus, it indicates that there are no statistically significant differences between values of reliability coefficients.

6. Conclusion

Through the previous sections, it is clear that the test's study achieved its goals by answering the questions and determining the effect of the length of the tests. The main results included the following:

The results of the study showed that the information function of the test differed with the variation models, as well as the absence of statistically significant differences attributed to the values of the empirical validity coefficient.Further studies along the tests are recommended to determine other effects.

References

Abu Allam, R.(2005). LearningEvaluation, Amman: Dar Al Masirah Printing & Publishing house

Abu Hatab, F. (1992). Teacher's Guide in Student's Assessment National Center for Examinations and Educational Assessment in cooperation with the Ministry of Education, Cairo: Dar Ghraib for Printing.

Adas, A. (2002). Teacher's Guide for Building Achievement Tests: Dar Al Fikr for Printing, Publishing and Distribution.

Al-Diabat, L. (2007). The effect of the length of the test on the characteristics of the distribution of real marks according to the logistic parameter model. An unpublished master thesis. Yarmouk University

Alhawari, A. (2015). The Effect of Test Length and Ability Distribution Form on the Estimation of A person's Ability According to Rasch Model in Item, An-Najah University Journal for Research: Humanities 29 (8), 1463 – 1488.

Alqudah, A.& Al-Shraifin, N. (2020). The effect of the length of a questionnaire on the accuracy estimations of the ability and the psychometric properties of the item and scale in the light of item response theory, An-Najah University Journal for Research: Humanities 34 (6), 953 – 982.

Allam, S.(1987). A critical balance study of the latent traits' models, and classical models of psychological and educational measurement. Kuwait University, The Arab Journal of Human Sciences, Issue (27), 18-44

Allam, S. (2005). Response models for Uni- dimensional and multidimensional test items and its applications in psychological and educational measurement. Cairo: Arab Thought House

Allen, D. (2016). The Impact of Shortening a Long Survey on Response Rate and Response Quality. Brigham Young University

Allen, M., & Yen, W. (1979). Introduction to Measurement theory. California: Cole Publishing company.

Al-Omri, H. (1985). Effect of the length of the test time on the performance of the examinees and on its psychometric characteristics among other subjects on risk degree and their anxiety level. A master degree unpublished thesis. Yarmouk University

Al-Qaisi, A. (2016). The Effects of the Sample Size and the Length of the Test on the Accuracy Estimation of the Item Parameters by Using Non-Parametric Item Response Theory, Mu'tah Research and Studies, Humanities and Social Sciences Series, (31)5, 203-246

Crocker, L., &Algina, J. (1986). Introductions to classical and modern Test Theory. Orlando, FL: Hacourt Brace Jovanovich.

Gronlund, N. (1985). Measurement and evaluation in teaching. New York: Macmillan.

Hakstain, A. & Whalen, T. (1976). A k-sample significance test for independent alpha coefficients. Psychometric, 41, 219-231.

Hambleton, R. & Swaminathan, H. (1985) Item Response Theory, Principles and Applications. Boston: Kluwer. Nijhoff Publishing. A member of the Kluwer Academic Publishers Group

Samara et al. (1989). Principles of Measurement and Evaluation in Education: Second Edition, Dar Al-Fikr: Amman

Author Information

Dr. Ahmad Abdel Hafed Alqraleh
Measurement and Evaluation,Administratorof
Marwad Primary School, Alkarak, Jordan