Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia
Social and Behavioral Sciences
ELSEVIER
Procedia - Social and Behavioral Sciences 29 (2011) 1359 – 1367

International Conference on Education and Educational Psychology (ICEEPSY 2011)

A Comparison the Information Functions of the Item and Test in One, Two and Three Parametric Model of the Item Response Theory (IRT)

Ali Moghadamzadeh [a,*], Keyvan Salehi [b], Ebrahim Khodaie [a,b]

[a]*PhD student of Educational Measurement, University of Allameh Tabatabaei, Tehran, Iran*
[b]*PhD student of Educational Measurement and Evaluation, University of Tehran, Tehran, Iran*
[a,b]*Assistant Professor in Social Statistics, National Organization for Educational Testing, Tehran, Iran*

Abstract

Present study sets out to compare test and item function in one-, two-, and three-parameter models in the item response theory (IRT). First, information function is delineated upon drawing on theoretical and psychometric foundations. The date collected from the pool of answers given to multiple-choice questions by candidates to the Iranian state universities who took the National Entrance Examination for State Universities in the Mathematics-Physics Testing Group in 2009 which was administered by the National Organization for Educational Testing (NOET). Systematic sampling was utilized to draw 2000 subjects from applicants who took entrance exam in the Mathematics-Physics Testing Group of National Entrance Examination for State Universities. To analyze the data obtained from experimental group, two statistical software viz. SPSS and BILOG were run. The results yielded indicate that the value of test and item awareness function in the two- parameter model is higher than values in the one-and three –parameter models .Furthermore, findings illustrate that the value of test and item awareness function in the one –parameter models is not significantly different from the corresponding value in the three –parameter model. This study has implications for educational measurement and testing procedures especially for standardized multiple-choice test.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Dr Zafer Bekirogullari.

Keywords: Item-Response Theory (IRT); Information function; national university entrance; one-, two-, and three-parameter models; information Values; measurement error; testee's ability; specific curve;

1. Introduction

Since test scores are used to make important decisions about testee's future in relation to their education, jobs etc, the effectiveness of tests is of prime importance (Thorandike, 1982). Having selected the test's latent trait model, the accuracy with which the test can estimate the testee's ability is to be determined. As every test is composed of a number of items each of which contributes to the accuracy of measurement, one can, through the information function of test items, gauge each test item information function at any given ability level, thereby determining the whole test information function (Hambelton & Cook, 1977).

* Corresponding author. Tel.: +989121944254; fax: +982188922237.
E-mail address: sanjeshali30@gmail.com.
This research was supported by the National Organization for Educational Testing (NOET) of Iran.

1877-0428 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Dr Zafer Bekirogullari.
doi:10.1016/j.sbspro.2011.11.374

The term “information” is used to mean that the testee possesses the required knowledge about the subject in focus. In statistics, this term has the same meaning, though a little bit less technical. The concept of information was subsumed in statistics by Fisher (1974) who defined it as the reverse of the accuracy index of estimated parameter (Baker, 2001).

If a parameter is accurately estimated, more information will be acquired about the numerical value of that parameter compared to when the same parameter is less accurately estimated. A parameter is measured through estimated variants based on the numerical value the parameter. Therefore, the accuracy level, which if different from the measures used, has formulaic expression. In fact, information value is obtained from formula 1 below.

$$I = \frac{1}{\sigma^2}$$

What is of significance to us is in the IRT theory is the estimation of a testee’s ability parameter (θ). A high information value indicates that the testee’s true ability, which is also high, can be estimated with a great degree of accuracy (Hambelton & Wanderlanden, 1982).

Information value can be estimated from -∞ to -∞ on the ability scale. As “ability” is a continuous scale, “information” is a continuous variable as well (Friedrich 2004). Of course, information value is expressed in the form of a distribution and the maximum distribution of information value reaches its height at one point (Springer, 1997).

In tests, the ideal information function is linear, enjoys a high value and estimates all ability levels with accuracy (Baker 2001). Based on formula 1, it can be contended that the information value at a certain point is the reverse of variance.

If information value is high, it can be concluded that a testee’s ability representing his true ability at that point can be estimated with accuracy; that is, all estimates will approach the true information value, however. If information value is low, the testee’s ability cannot be estimated with accuracy and estimation values will basically be less evenly distributed around the true ability level (Friedrich, 2004).

Aim of present study was to investigation of information function in 1, 2, and 3 IRT model based on Iranian national university entrance exam data base. Not only information function of item and test was described, but also comparison of information function in different ability levels in three IRT model and its relation with measurement error was examined.

2. Theoretical framework

2-1- Information function of test items

The information value pertaining to each item for each ability level can be obtained through formula 2 below.

$$I_i(\theta) = \frac{(P'_i(\theta))^2}{P_i(\theta)Q_i(\theta)}$$

The information function of an item for a given ability level can be defined as the proportion of the square root of the differentiation of item characteristics to its variance (Hambelton, 1989). The information function represented in formula 2 has a normal ogive form. Instead of this function the logical function, which yields a little bit different results, can be utilized (Lord, 1980).

Suggested formula 3 below for the estimation of the information value of an item

$$I_i(\theta) = \frac{P^2 a_i^2 (1 - c_i)}{(c_i + e^{DL_i})(1 + e^{-DL_i})^2}$$

In this formula a_i and b_i and c_i represents parameters of discrimination, difficulty, and guessing, respectively. In this formula both the slop and the variance of the scores of the item depend on the amount of item parameters questions with steeper slopes a point with maximum slop-produce more information value in regard to the testee’s ability. The difficulty parameter also shows the maximum likelihood of the information function of items on the ability continuum. Items whose slopes are steeper provide researchers, teachers and decision-makers with more information compared to items whose slopes are less steep. In fact, whenever, the ration of two parameters is 2 to 2, the resultant maximum information would be 4 to 1.

When guessing is possible, the information would be uneven because weak testees resort to guessing more often/ leading to a decrease in the information value of the item on the left side of the scale (Lord, 1980). When the slope is steep and the variance is low, information function would be larger; however, when the slope is not that steep and variance is great, information function would be less (Alan & Yen, 1979).

2-2- Test information function

Since it is all the test items in a test that are used to estimate a testee's ability, the information value obtained through the test can be calculated for every ability level. A test is a combination of items; hence, test information function can, for any given ability level, be simply obtained from the sum of the information function of all the items. This means that test information function can be expressed as follows:

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta)$$

$I(\theta)$: Value information test

$(\theta)\ I_i$: Value information item

The Information function of the whole test will be much greater than that of the individual items. Consequently, a complete test measures people's ability more accurately and more comprehensively than each individual item. The important point inherent in formula 4 is that the very existence of more items in a test increases the information function of the test. Therefore, we can generally contend that longer tests measure a testee's ability more accurately than shorter ones do (Baker, 2001).

The use of tests for other specific purposes requires the use of other forms of information function. Although it is possible to obtain the information function of each item, this is rarely done. The information value gained form each item is negligible, so the information value gained from the whole test is of prime importance. That is why the testee's information value at one ability level and its information function are headed. Of course, since a test's information obtained through the sum of all he items' information value at a given level, the item information value should first be defined. It is to be noted that the mathematical definition of the item information value depends on the specific curve used (Friedrich, 1997).

2-3- Curve Model of One-Parameter Item

In the one-parameter model, item information value is defined as:

$$I_i(\theta) = P_i(\theta)Q_i(\theta)$$

Information functions play a significant role in the IRT because information value depends on item difficultly (Friendrich, 2004). In the one-parameter model the numerical value of item discrimination is assumed to be one. The possibility of guessing is zero in this model, resulting in a correspondence between item difficulty and the testee's ability (Hambelton & Jones, 1993).

2-4-Curve Model of Two-Parameter Items

In a two-parameter model, information function is defined as:

$$I_i(\theta) = a_i^{2} P_i(\theta)Q_i(\theta)$$

$$\mathrm{Q_i}(\theta) = 1 - \mathrm{P_i}(\theta)$$

In this formula a_i stands for discrimination power parameter and the value of P_iQ_i is calculated through the following formula:

$$P_i(\theta) = 1/(1 + EXP(-a_i(\theta - b_i))$$

$$Q_i(\theta) = 1 - P_i(\theta)$$

2-5- Curve Model of Three-Parameter Items

This model does not enjoy the same features as logistic functions do. Hence, to calculate item information value the following complicated formula is utilized.

$$I_i(\theta) = a^2 \left[\frac{Q_i(\theta)}{P_i(\theta)} \right] \left[\frac{P_i(\theta) - c^2}{1 - c^2} \right]$$

$$P_i(\theta) = c + (1 - c)(1/(1 + EXP(-l)$$

$$L = a(\theta - b)$$

As the guessing effect does exist in the three-parameter model, the information value in this model is less than such value in the two-parameter model when **a** and **b** values are the same. This means that parameter model better calculates item information value that the three-parameter model (Baker, 2004).

Since the form of a test's information function depends on the purpose for which the test is designed, there can be different general interpretations. The test information function, which rises dramatically at some points on the ability scale measures on ability along the ability scale with different degrees of accuracy. Such tests work better when estimating the ability of those testees whose ability is near or at the top of the ability scale. In some tests, information function is more even in some areas on the ability scale. Such tests estimates testee's ability in a certain area with almost the same accuracy, but the accuracy fluctuates when the test functions outside, that certain area. Therefore, when interpreting the results, one should consider those testees whose ability fall within a certain area. And it is more important to note that the reverse relationship between information value and the variability of ability estimation to change information value to the standard error estimation, which requires that one obtains the square root of the test's information value.

In general terms, it can be stated that in the new measurement models, the item information function may be used to gain information about the contribution of each question to the determination of testee's trait level and determine, by way of local independence, the contribution of each item to the effectiveness of the whole test without regard to the other items in the test. It is also possible to use the test's information function at a certain ability level to compare and evaluate different tests. However, in the classical model of testing there is no an appropriate method to determine the information value each items provides about testee's ability. In addition, in the classical model the contribution of each item to the testee's reliability and validity depends on the characteristics of the other test items; that is, the contribution of each item to the test's reliability and validity can not be determined. In fact, test validity is a function of all the items in the test and the contribution of each item to test reliability cannot be determined independent of the other test items (Lord, 1975, 1980).

One of the other applications of item and test information function is its use to design tests with specific purposes and select the required questions. To this end, the intended items given to a number of testees; then the latent information trait of each item is calculated based on the model under discussion. To design a test for a specific purpose the following steps are, as recommended by (Lord, 1980), taken:

1. The form of the test information curve is determined; that is, the intended level on the ability continuum is decided upon. Then the form of the test information curved is determined based on the purpose of the test which specifies the required information value at a specific ability level.
2. Based on each item's information value, the items are selected in such a way that they cover the area below the information curve.
3. The information functions of the items are cumulatively added up to obtain a satisfactory estimation of the information curve.

The test information function can be used to construct parallel tests as well as those for specific purposes. That is, information function can be utilized to construct both weak and strong tests. Two forms of a test one considered as parallel when they measure the same ability and possess similar information function. Strong parallel forms

should, in addition to the already maintained conditions, fulfill another condition to the effect that the item parameters should be equal in both forms.
To compare the effectiveness of different tests is another application of test information function. Majority of researchers inclined towards comparing information values obtained from different testing patterns or scoring methods of a test. For example, you may need to compare the information function of several standard tests to find out which test provides more information in an area or areas of interest. In such cases, the information functions of tests are calculated; the index so obtained is known as the relative effectiveness index.

$$RE(x_1, x_2, \theta) = \frac{I(x_1, \theta)}{I(x_2, \theta)}$$

In this formula the relative accuracy is indicative of test information function (Alan & Yen, 1979).

3. Methodology

3-1- Population and Sampling Method

To answer the intended items, the subjects for this population were considered. The people were all Iranian and included all the 2009 university entrance exam applicants who were taking the test in the field of Maths.

Table 1: Number of applicants of the 2004 university entrance exam in the field Maths.

Field	Female	Percentage	Male	Percentage	Total number
Maths	138886	45.91	163643	54.02	302511

The sampling method was a regular one based on which 2000 people from among the list of all applicants were randomly selected. Of course, those who had obtained one-fourth of the total score were not included in this study to control for the role of the guess factor. In the IRT model researchers are advised to use 200 people for one-parametric models 500 for two-parameter models (wright & Stone, 1979), and over 1000 for three-parameter models (Tissen & Vainer 1982). Hambelton (1989) also agrees that a sample of at least 1000 people is required in three-parametric models. Hence, for the three-parameter of this study a sample of 2000 subjects were selected.

3-2- Instrumentation and Procedures

The university entrance exam (UEE) in Iran is an achievement exam. The exam is prepared by a group of experienced and expert test designers at the National Organization for Educational Testing (NOET). What was used in this research to gather data was part of the answer sheets of the UEE applicants related to the math section of the exam which included 55 test items. It is to be noted that experts contend that the minimum number of items to be investigated for a study like the present one is to be 20 and that the number of items in the subtests is affected by the number of the parameters in the model used. Hence, it is often argued that the adequate number of items for one-parameter, two-parameter and three-parameter models of subjects is large, acceptable estimations can be made even when the number of items is 20 (Hambleton, 1989).

4. Results and Findings

To analyze the data, use was made of the BILOG computer software and the information function of items was calculated for the three models with due consideration to the ability level. Through the BILOG software the parameters of the items for the three models were calculated. The information value was calculated for the whole exam with regard to the one-, two-, and three-parameter models.
As table 2 demonstrates the information value was calculated for the whole exam with regard to the one-, two-, and three-parameter models.

Table 2: information values in the one-, two-, and three-parameter models

Question	Information function		
	one-parametric	Two-parametric	Three-parametric
1	0.84	1.12	0.74
2	0.15	0.74	0.13
3	0.87	1.20	0.88
4	1.74	2.50	1.90
5	0.25	0.37	0.20
6	0.91	1.20	0.94
7	0.97	1.25	0.89
8	0.21	0.62	0.28
9	0.73	0.94	0.68
10	0.23	0.57	0.21
11	0.49	0.74	0.40
12	0.09	0.20	0.11
13	0.35	0.69	0.47
14	0.29	0.61	0.20
15	0.94	1.24	0.90
16	0.37	0.84	0.30
17	1.72	1.28	1.80
18	0.93	1.08	0.89
19	0.83	0.91	0.80
20	0.73	0.90	0.89
21	0.73	0.89	0.80
22	1.25	2.14	1.65
23	0.73	0.49	0.54
24	0.13	0.90	0.18
25	0.31	0.91	0.38
26	0.35	0.94	0.30
27	0.88	1.13	0.87
28	1.24	2.18	1.01
29	1.60	2.35	1.21
30	0.75	1.09	0.70
31	0.67	0.98	0.63
32	0.66	0.97	0.62
33	0.68	0.98	0.69
34	1.11	1.90	1.18
35	0.98	1.34	0.90
36	0.40	0.85	0.41
37	0.82	1.17	0.80
38	0.33	0.69	0.30
39	0.92	1.13	0.90
40	0.14	0.29	0.11
41	1.50	2.64	1.40
42	0.92	1.17	0.20
43	0.82	1.03	0.80
44	0.40	0.90	0.39
45	0.86	1.04	0.98
46	0.32	0.54	0.48

Table 3: comparison of items information function based on ability level in one-parameter model

θ	-3	-2	-1	0	+1	+2	+3
$I_i(\theta)$	0.05	0.11	0.26	1.19	1.24	1.21	0.28

As table 3 reveals the information function yields the highest estimation at ability level 1 and the lowest the ability level 3. On other words the higher the information value, the less the standard error of measurement.

As the observed values indicate, the question information function increases steadily. This increase reaches its maximum of 1.24 at the ability level 1 of and then begins to decrease after this ability level; in a sense, it can be stated that the two sides of the ability level of 1 or the information function of 1.24 are to some extent symmetrical.

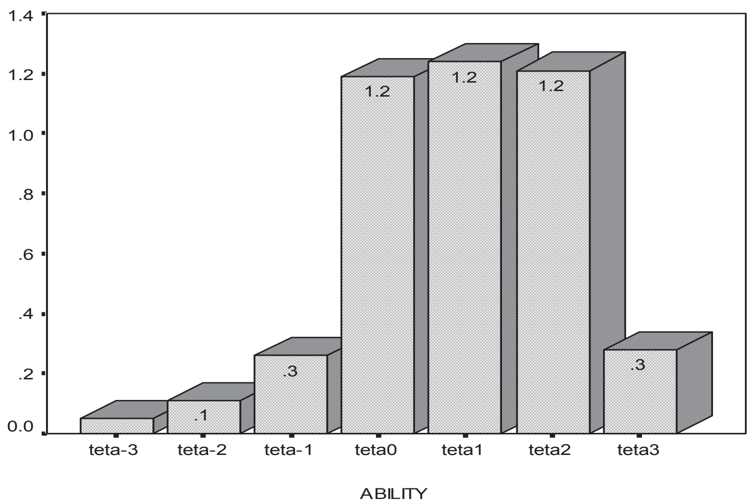

Figure 1: Test information function in one-parameter model

Table 4: comparison of item information function based on ability level in two-parameter model

θ	-3	-2	-1	0	+1	+2	+3
$I_i(\theta)$	0.02	0.06	0.40	1.36	2.67	1.45	0.38

As table 4 reveals the information function has been calculated based on the two-parameter model. While item difficulty in the one-parameter model is 2, it is 1.5 in the two-parameter model and the information value has been calculated in terms of the ability level. The results obtained indicate that in this model the ability level of 1 offers the highest estimation of the information function and the lowest estimation relates to the ability level of 3. In this model the information function increases steadily too, which is in harmony with the increase in the ability level. However, the highest value of the information function is at the ability level of 1 form which point onwards the information function begins to decrease.

When comparing the two-parameter model with the one-parameter model, it can be concluded that the information function is higher at all ability levels than it is in the one-parameter model; this is because of the discrimination index which is assumed to be 1.5 in the two-parameter model while it is assumed to be a fixed coefficient.

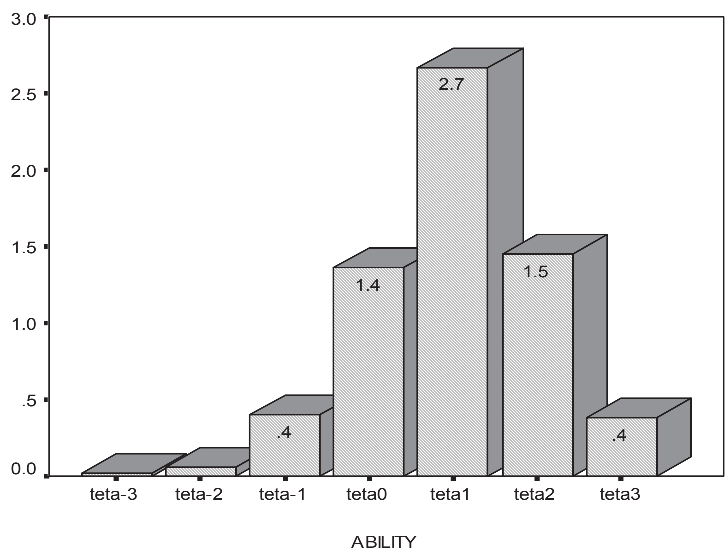

Figure 2: Test information function in the two-parameter model

Table 5: comparison of item information function based on ability level in three-parameter model

θ	-3	-2	-1	0	+1	+2	+3
$I_i(\theta)$	0.01	0.07	0.48	1.24	1.78	1.31	0.45

As table 5 illustrates the information function has been calculated based on the three-parameter model for different ability levels. In table 6, the ability level of 1 has the highest information function value. The increase of the information function is steady in this model too. And the highest information value, which is 1.78, is at the ability level 1 from which point onwards the value of the information function begins to decrease. A look at the two sides of the information function at the ability level of 1 shows that a symmetrical curve has been formed.

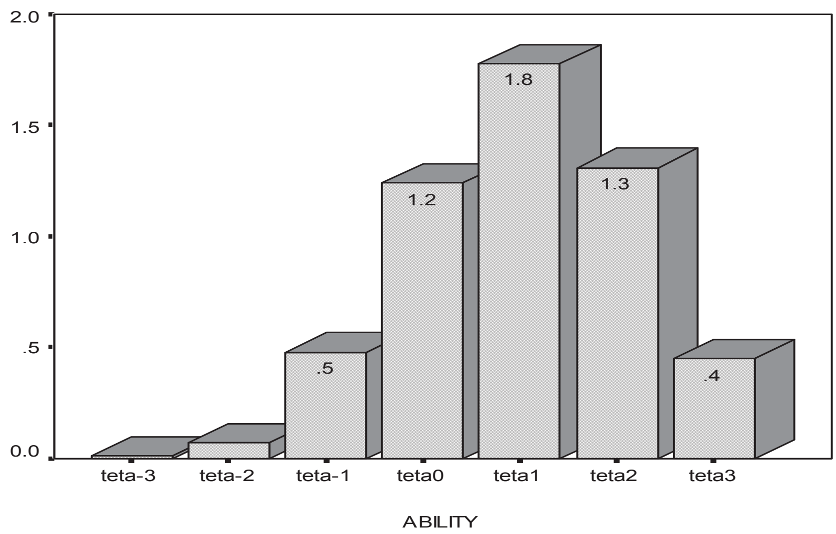

Figure 3: Test information function in three-parameter model

5. Conclusion and discussion

In IRT there is reverse relationship between the information function and the standard error of measurement. It means that the higher the numerical value of the information function is, the less likely errors emerge.

With consideration to the estimated values in the three models, it can be concluded that the estimated values of the information function in the one-parameter model are less than the ones in the two-parameter one and that the estimated values of the information function in the two-parameter model higher than the ones in the three-parameter one. However, there is no much difference in the information values in the one- and two-parameter models, which is in accord with the studies conducted by Hambelton (1977) and Friedrich (2004). On the whole, it can be concluded that in the two-parameter model the guessing factors are eliminated; that is, when items involve a high index of guessing the likelihood of error increases. That is why it can be stated that there is a reverse relationship between the information function and the standard error of measurement. In another study, Hambelton (1989) investigated the accuracy of estimation in the light of the guessing factor and found out that the guessing factor increases measurement error. It has to be remembered that Brinbaum (1986) and Divgi (1986) studied the effect of guessing in the three models and reached similar conclusions.

The present study revealed that the reliability indices calculated based on ability levels are different. In IRT it is actually possible to construct a test that exactly corresponds with the objectives of the test designers or its use. And to determine the accuracy of the test, the information function can be utilized. That is why in IRT there is more control over the preparation of tests for specific purposes (Lord, 1980). Since in the classical model the reliability of the whole test is calculated, the contribution of each test item to reliability is not independent of that of the other test items. However, in IRT the information function is calculated for each individual test item, the contribution of each item to reliability is calculated without regard to the other test items, and the contribution of each item to the overall effectiveness of the test is determined. One of the other applications of the test and item information function is its use to design tests for specific purpose.

The results obtained in this research study also confirmed the points discussed above and revealed that the information value is different in the IRT three models and that there is a reverse relationship between the information value and the standard error of measurement (Parand, 2002).

Acknowledgment

I am heartily thankful to Dr. Falsafinejad, whose encouragement and guidance from the initial to the final level enabled me to develop an understanding of the subject.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Cole Publishing Company, Monterey, California, a division of Wadsworth, Inc.

Baker, Frank. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Baker, F. (2004). *Item Response Theory: Parameter Estimation Techniques*, Second Edition (Statistics, a Series of Textbooks and Monographs) CRC.

Brinbaum, A. (1986). *Some Latent trait models and their use in inferring an examinee's ability.* In F.M. Lord & M.R. Novick, Statistical theories of mental test score. Ma: Addison – Wesley.

Divgi, D. R. (1986). *Does the Rash model really work for multiple choice items? Not of you look closely. Journal of Educational Measurement.* 23, P: 283-298**.**

Freidrich, S. (2004). *A visual guide to item response theory*. Ivailo Partchev

Hambleton, R.K. (1989). *Principles and Selected applications of Item Response Theory.* InR. Linn (ed), Educational Measurement (3rd end). New York: Memillan. P: 147-200.

Hambleton, R.K., & Cook, L.L. (1977). *Latent trait models and their use in the analysis of educational test data.* Journal of Educational Measurement. 14(2). P: 75-94.

Hambleton, R.K., & Jones, R.W. (1993). *An NCME Instructional Model on comparison of classical test theory and Item response theory and their applications to test Development*. Educational Measurement: Issues and Practice. P: 38-46.

Hambleton, R.K., & Vander L., Wim. J. (1982). *Advance in item response Theory and Applications: An introduction.* Applied Psychological Measurement. 6(4). P: 373-378.

Lord, F.M. (1975). The "ability" Scale in item characteristic curve theory Psychometrika. 40, P: 205-217.

Lord, F.M. (1980). *Applications of item Response theory to practical Testing Problems,* Hillsdale, N.G.: Lawrence Ertaun, 14, P: 117-138.

Parand, A. (2003). Construct Parallel tests using classic and item response Theory (IRT) methods. . *MA Thesis*, Alameh University. Tehran, Iran.

Springer, F. (1997). *Weakly Parrallel test in latent trait theory with some criticisms of classical test theory*. Psychometrika. 42. P: 193-198.

Thissen, D., & Wainer, H. (1982). *Some Standard errors in item response theory,* Psychometrika, 47, P: 397-412.

Thorndike, R. (1982). *Applied Psychology*.Houghton Mifflin Company Boston Dalas Geneva, Illinois Hopewell, New Jersey Palo Alto London.

Wright, B.D., & Stone, M.H., (1979). *Best test design.* Chicago: MESA Press.