GU J Sci 30(1): 235-249(2017)

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

An Evaluation of the Two Parameter (2-PL) IRT Models Through a Simulation Study

Hülya OLMUŞ[1], Ezgi NAZMAN[1,*], Semra ERBAŞ[1]

[1]*Department of Statistics, Faculty of Science, University of Gazi, Beşevler-Teknikokullar, Ankara. TURKEY*

Article Info

Received: 15/08/2016
Accepted: 02/11/2016

Keywords

Two-parameter item response theory
Joint maximum likelihood estimation
Parameter estimation
Discrimination parameter
Ability parameter

Abstract

The aim of the study is to evaluate parameter estimation of two-parameter item response theory (2-PL IRT model) using Joint Maximum Likelihood (JML) estimation technique. Hence, a simulation study is conducted in terms of various sample sizes, item numbers and ability parameter levels for low and high discirimination parameter levels. Hereby, the examinee ability parameter, item discrimination parameter and difficulty parameter are obtained as well as Test Information Function (TIF) and point-biserial correlation. One of the highlighted results indicates that the level of discrimination parameter plays an important role in parameter estimation for 2-PL IRT models.

1. INTRODUCTION

The progress in science is mostly possible with the help of some measurement techniques. However, evaluating results of measurement techniques is difficult in some fields where unobserved variables being assumed as latent variable of the examinee (person) play a significant role. Mathematical or verbal ability of students in education, consumer preferences in marketing, political attitude of voters in politics…etc. can be given as example of the field where latent variable is not able to be directly measured by the conventional measurement methods (Rasch, 1960; Rizopoulos, 2006).

Item response theory (IRT) provides an acceptable framework to measure the amount of latent variable that examinee has. The measurement process often starts with using a questionnarie or an examination. Items or test questions has always been crucial in both measurement and evaluation steps in a research. In IRT, item responses are outcomes (dependent) variables and the ability of examinee and characteristics of items are latent predictors (independent) variables (Le, 2013). Even though estimation of the examinee latent variable is the primary aim of testing, this aim is only possible with determining item parameters. Thus, IRT is a helpful method from the point of estimation both examinee parameter (latent variable) and item parameters (Birnbaum, 1968).

IRT became popular in the last three decades, however the usage of IRT has been examined in detail for almost a century. The fundamental reason of being a young known method is due to the comprehensive computational necessities of the IRT method. Until 1980's, Classical Test Theory (CTT) had been preferred especially in education and psychology. In the last part of twentieth century, IRT gained applicability with the innovation in computer technology. Nowadays, the importance of IRT usage and the development of the model has been widely increasing in many fields. The detailed information can be found in the book of Andersen, 1977; Hambleton, 1990; Baker, 2001; Baker&Kim, 2004.

Unlike CTT, IRT method does not deal with sum of correct score numbers in order to evaluate an examinee's performance. On the contrary, IRT assumes the contribution equality of the items to the overall

*Corresponding author, e-mail: ezgicabuk@gazi.edu.tr

scores. Moreover, IRT assumes that items can vary in their difficulty level while examinees can vary in their latent trait level. Since examinees with the same sum of score may differ in their trait measurement, IRT may give more accurate results with regards to the latent traits of examinees. Both comparison and estimation for examinee and item parameters are possible with the help of IRT (Harris, 1989).

Nowadays, measurement and evaluation process has been taking on a new shape with the help of developing more exploratory measurement methods. One of the developing methods providing estimation of both item and examinee (person) parameters is IRT. Dichotomous IRT, which is one of the IRT models, has been popular for last a few decades. This model is a special form of Generalized Linear Models (GLM) and has a significant importance. In this study, two parameter (2-PL) IRT, which is one of the dichotomous IRT models, is examined. This model is studied in order to evaluate parameter estimations in terms of various sample sizes (n), item numbers (k) and ability parameter levels (θ) for two different discrimination parameter levels (a). The aim of the study is to reveal the effects of different sample sizes, item numbers and ability levels on the estimation ability and difficulty parameters as well as test information function and point-biserial correlation. In addition, low and high level item discrimination level (a) are studied separately to understand how item discrimination level is efficient on obtained results.

In accordance with this purpose, the study is organized as follows: The two parameter IRT model, one of dichotomous IRT models, is explained in section 2. Parameter estimation technique using in the study is mentioned in section3. Test information function and point-biserial correlation are explained in section 4. Simulation study and obtained results are shown in section 5. A numerical example is given in section 6. The outstanding findings are evaluated in section 7.

2. A GENERAL VIEW OF ITEM RESPONSE THEORY

In IRT models, the three most commonly used models are one parameter (1-PL), two parameter (2-PL) and three parameter (3-PL) IRT models. 1-PL IRT model contains only discrimination parameter whereas 2-PL IRT model contains both discrimination (a) and difficulty parameters (b). On the other hand, 3-PL IRT model is consist of the psedue-chance parameter (c) in addition to the discrimination and difficulty parameters (Andersen, 1977;Hambleton, 1990).

Item discrimination parameter ranges from $-\infty$ to $+\infty$. If an item has the highest discrimination value, it means that item may discriminate examinees more accurately and clearly. Item difficulty parameter which ranges from $-\infty$ to $+\infty$, but in practice, b values generally ranges from -3 to +3, when the examinee parameter (θ) has standart normal distribution. If an item has the highest b value, it refers to the hardest items. Therefore, the probability of correct response for examinee having low ability will be also low. Also, examinee parameter ranges from $-\infty$ to $+\infty$, and in practice θ values ranges from -3 to +3 (Mellenberg, 1994;Baker, 2001).

IRT models have three assumptions that should be satisfied. The first of these is the unidimensionality assumption implying that all items should measure only one examinee parameter. The second assumption is the local independence stating that an examinee's correct response is only based on the latent variable of the examinee. The last assumption is the monotonicity implying there exist a monotonic non-decreasing relationship between latent variable of examinee and probability of answering correctly a certain item (Paolino, 2013). Although there exist three basic assumptions, some studies shows that IRT models are robust for the situations that assumptions are violated (Hulin et al., 1983).

There are two commonly used link functions in IRT models: Logit and Probit. On the other hand, Logit is the most widely preferrred link function because of the easiness in the computation of parameter estimation (Mellenberg, 1994;Baur&Lukes, 2009).

2.1. Two Parameter (2-PL) IRT Model

The 2-PL IRT model, which was proposed by Birnbaum, 1968, includes item discrimination parameter in the model. This model is also able to be obtained from 3-PL IRT model when the pseudo-chance parameter is assumed zero.

The 2-PL IRT model is expressed in the Equation 1.

$$P(Y_{ij}=1|\theta_i)=\frac{1}{1+\exp(a_j\theta_i-b_j)} \text{ for } i=1,\dots,n \text{ and } j=1,\dots,k \quad (1)$$

where

θ_i : ability parameter of *i* th examinee

a_j : discrimination parameter of *j* th item

b_j : difficulty parameter of *j* th item

n :number of examinee

k :number of item

Logit link function is $(a_j\theta_i-b_j)$ in Equation (1) that shows the probability that a particular examinee with latent ability score of θ_i correctly answers item *j* (Hulin et al.,1983).

3. PARAMETER ESTIMATION OF 2-PL IRT MODELS

Estimation of both item and examinee parameters is a crucial process in IRT. The most widely used techniques in estimation of the IRT models are Conditional Maximum Likelihood, Marginal Maximum Likelihood, Joint Maximum Likelihood, and Bayesian Maximum Likelihood. Each parameter estimation technique has its own special properties and limitations. All the techniques contain a likelihood funtion and distribution function, so the main idea behind these tecnhiques is the maximization of the likelihood function.

In the context of IRT, a likelihood function can be the probability of observing a particular pattern of responses from an individual, or it can be the probability of observing a particular response matrix.

Let $y_{i1},y_{i2},...,y_{ik}$ be the dichotomous response of the *i* th examinee to *k* test items, $\mathbf{a}=(a_1,...,a_k)$ and $\mathbf{b}=(b_1,b_2,...,b_k)$ be the vectors of discrimination and difficulty parameters, respectively. When it is assumed that an examinee taking test responses each item independently, and then the probability of observing the all placement of the *i* th examinee is given as:

$$P(Y_{i1}=y_{i1},...,Y_{ik}=y\setminus\theta_i,\mathbf{a},\mathbf{b})=\prod_{j=1}^{k}P(Y_{ij}=y_{ij}\setminus\theta_i,\mathbf{a},\mathbf{b}) \quad (2)$$

Then the likelihood function for all responses of examinees is shown as:

$$L(\boldsymbol{\theta},\mathbf{a},\mathbf{b})=\prod_{i=1}^{n}P_j^{y_{ij}}\left(1-P_j\right)^{1-y_{ij}} \text{ for } i=1,\dots,\text{n and } j=1,\dots,\text{k} \quad (3)$$

where

n : number of examinee

k : number of item

and the full log-likelihood for examinees is shown in Equation (4).

$$\ln L=\prod_{i=1}^{n}\prod_{j=1}^{k}\left[y_{ij}\ln(P_j)+(1-y_{ij})\ln(1-P_j)\right] \quad (4)$$

In this study, Joint Maximum Likelihood (JML) technique was used to estimate both item and examinee parameters treating parameters as fixed. Then, the process yields estimations for both item and examinee parameters. Fundamentally, the technique of JML estimation is based on logistic regression with dummy variables for the item parameters and the examinee abilities.

The estimations of parameters simultaneously are only possible with an iteration process due to the non-linearity of the first and second derivations of the parameters. In literature, Newton-Raphson, Fisher's Method of Scoring and Expectation Maximization are given as some of the recommended methods (Toribio, 2006). In the study, Newton Raphson method was applied to estimate both item and person parameters.

The steps of parameter estimation for the JML technique starts with initial item and examinee parameters treating as known to estimate the item parameters via Newton-Raphson method. This method proceeds until the difference between successive iterations become sufficiently small (Baker&Kim, 2004; Cai&Thissen, 2014).

4. SOME DESCRIPTIVE STATISTICS FOR IRT

In this section, Test Information Function (TIF) and point-biserial correlation ($\boldsymbol{r_{pbi}}$), which are two important concepts, are comprehensively given. It is benefited from the TIF to reveal the information how much ability each examinee has and the strong effect of the ability parameter on the discrimination parameter. Since it shows whether discrimination of the question is high or low, point-biserial correlation provides an important contribution to the researchers. It gives an objective point of view to the researchers on the determination which questions should be prepeared in the future.

4.1. Test Information Function (TIF)

In IRT, the general interest is the estimated value of ability parameter for an examinee. The amount of information based on an item is able to be computed for any ability level. Item Information Function (IIF) is shown as $I_i(\theta)$ $i = 1,\dots,n$ where *n* is the number of examinees. 2-PL IIF is shown as:

$$I(\theta,a,b) = a^2 P(\theta) Q(\theta) \tag{5}$$

As it is clearly seen in the Equation (5), discrimination parameter value has importance in computing item information function.

A study is a set of items, so the test information for the given ability level is computed from the sum of the item informations at that level. TIF is defined as:

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta) \tag{6}$$

In general, the TIF will be higher than that for a single IIF. Therefore, a test or questionnaire is able to measure more precisely than what a single item indicates (Baker, 2001).

4.2. Point-biserial Correlation $(\boldsymbol{r_{pbi}})$

Point-biserial correlation is a statistical coefficient using for estimation the degree of relationship between an item and the total score. It has a naturally dichotomous nominal scale and another has an interval or ratio scale. The point-biserial correlation in IRT is shown as:

$$r_{pbi} = \frac{M_p - M_q}{S_t}\sqrt{pq} \tag{7}$$

where

M_p : whole test mean for examinees responding item correctly

M_q : whole test mean for examinees responding item incorrectly

St : standard deviation for whole test

p : proportion of students responding correctly

q : proportion of students responding incorrectly

Point-biserial correlation ranges between -1 and +1. The high point-biserial correlation value indicates strong relationship between two variables as is Pearson correlation coefficient (Baker&Kim, 2004; Toribio, 2006).

5. SIMULATION STUDY

The aim of our simulation study is to make comparison of estimated item and examinee parameters as well as point-biserial correlation and TIF. Therefore, three different ability levels were chosen in order to observe the changes with regards to low, neutral and high ability levels by using MATLAB software. In the simulation study, different item numbers (k=20,40,60,90,120,150) and sample sizes (n=100,250,500,1000,5000) are determined for three different ability levels. Besides, item discrimination parameter is separately studied in the situation where discrimination parameter is both high and low. Once, the initial parameter values are specified in order to conduct the simulation study. Then, the examinees with low ability are generated from Normal distribution with a mean of -1 and variance of 1 as shown $N(-1,1)$. The examinees with high ability are generated from $N(1,1)$. The ability levels of examinees indicating there is not information are generated from $N(0,1)$. Besides, the difficulty parameters are generated from $N(0,1)$. The high discrimination parameter are generated from Uniform (0, 2) while the low discrimination parameters are generated from Uniform (0, 0.2).

Joint Maximum Likelihood Estimation technique was used to estimate parameters of 2-PL IRT model. The simulated nxk response matrix is generated from Bernoulli distribution. The probability of answering to a particular item correctly by a certain simulated examinee is conducted with the help of Logit link function when the values of initial parameters are determined.

Once the ability, difficulty and discrimination parameters are generated, then parameter estimations are calculated using JML estimation technique. The mean of obtained parameter estimations, point-biserial correlations and TIF for all the different situations is given in the Table 1 and Table 2, separately. The situation where the item discrimination parameters are both low and high are studied along with three different situations of ability level. The situation in which item discrimination parameter is low is shown in the Table 1 and high item discrimination parameter is also shown in the Table 2.

Table 1. *The results of parameter estimations, point-biserial correlation and test information function for low level di*

$a \sim Uniform(0, 0.2)$ $b \sim N(0,1)$		$\theta \sim N$ (-1,1)				$\theta \sim N$ (0,1)				
n	***k***	***a***	***b***	r_{pbi}	***TIF***	***a***	***b***	r_{pbi}	***TIF***	***a***
100	**20**	0.5596	0.4485	0.2247	1.4685	0.6071	-0.3146	0.2409	1.6897	0.55
	40	0.3693	0.1679	0.1561	1.7375	0.3605	-0.0321	0.1451	1.4755	0.40
	60	0.3447	0.1231	0.1417	2.1850	0.3073	-0.0020	0.1368	1.7637	0.32
	90	0.2910	0.2977	0.1220	2.2506	0.2842	-0.0318	0.1222	2.3066	0.27
	120	0.2340	-0.0443	0.0978	2.5521	0.2445	-0.0025	0.1049	2.5618	0.22
	150	0.2349	-0.0716	0.0999	2.9964	0.1971	0.0181	0.0868	2.7642	0.21
250	**20**	0.5684	-0.0024	0.2283	1.2890	0.6437	0.2522	0.2533	1.4980	0.54
	40	0.3738	0.1806	0.1611	1.3079	0.389	-0.0610	0.1668	1.3818	0.30
	60	0.2956	0.0672	0.1331	1.3460	0.335	-0.0484	0.1441	1.5295	0.27
	90	0.2511	0.1069	0.1084	1.4665	0.2549	-0.1308	0.1120	1.6019	0.24
	120	0.2293	0.1068	0.0998	1.7670	0.2278	-0.0213	0.1004	1.7895	0.25
	150	0.2106	0.1675	0.0921	1.9589	0.2091	-0.0041	0.092	1.9086	0.21
500	**20**	0.5115	0.2025	0.2231	1.1059	0.5280	0.2620	0.2279	1.1913	0.51
	40	0.3728	0.0116	0.1665	1.2602	0.3777	-0.1402	0.1620	1.2030	0.38
	60	0.2902	0.1755	0.1299	1.1717	0.3174	0.1717	0.1416	1.3463	0.30
	90	0.2555	0.2752	0.1127	1.3769	0.2470	0.0326	0.1095	1.2607	0.27
	120	0.2334	0.1401	0.1032	1.6176	0.2360	0.0299	0.1052	1.7221	0.22
	150	0.2052	0.0751	0.0920	1.6039	0.2084	-0.1734	0.0937	1.6861	0.20
1000	**20**	0.5348	0.1080	0.2321	1.1836	0.5610	0.0428	0.2261	1.1508	0.51
	40	0.3837	0.1403	0.1633	1.1536	0.3678	-0.0851	0.1630	1.1568	0.37
	60	0.3050	0.2091	0.1321	1.2073	0.3181	-0.2664	0.1383	1.2765	0.30
	90	0.2593	0.0991	0.1156	1.3646	0.2542	0.1539	0.1122	1.2580	0.24
	120	0.2293	0.1769	0.1015	1.4121	0.2396	-0.0479	0.1059	1.5633	0.22
	150	0.204	0.1568	0.0909	1.4609	0.1982	-0.0326	0.0878	1.3735	0.20
5000	**20**	0.5245	0.1584	0.2265	1.1123	0.5721	0.0488	0.2258	1.1327	0.53
	40	0.3874	0.4201	0.1621	1.1237	0.3626	-0.0082	0.1588	1.0660	0.37
	60	0.3044	0.0874	0.1369	1.1789	0.3144	-0.1706	0.1356	1.1841	0.31
	90	0.2524	0.1432	0.1123	1.2075	0.2565	-0.1167	0.1147	1.2422	0.20
	120	0.2308	0.1681	0.1031	1.3609	0.2277	0.0241	0.1018	1.3242	0.22
	150	0.2125	0.1633	0.0948	1.4385	0.2080	0.0808	0.0923	1.3730	0.20

Table 2. *The results of parameter estimations, point-biserial correlation and test information function for high level d*

$a \sim Uniform(0,2)$ $b \sim N(0,1)$		$\theta \sim N$ (-1,1)				$\theta \sim N$ (0,1)				
n	***k***	***a***	***b***	r_{pbi}	***TIF***	***a***	***b***	r_{pbi}	***TIF***	***a***
100	**20**	1.2622	1.0708	0.4032	5.8999	1.2806	-0.0501	0.4138	5.4621	1.16
	40	1.3717	1.2761	0.3974	11.8134	1.3534	-0.1126	0.4342	13.8861	1.36
	60	1.4977	1.5855	0.4079	17.1148	1.4638	-0.0928	0.4020	27.0956	1.16
	90	1.4259	1.1857	0.4137	29.1952	1.2162	-0.0410	0.3528	29.1702	1.18
	120	1.4101	0.8929	0.4373	39.5551	1.2044	-0.2442	0.3956	34.5469	1.17
	150	1.3588	1.1772	0.3730	46.0421	1.0098	-0.1822	0.3417	35.0662	1.04
250	**20**	1.2566	1.0835	0.3542	5.4833	1.3164	-0.2936	0.4261	6.6538	1.08
	40	1.2875	0.7315	0.3785	12.1460	1.2875	0.1797	0.4229	13.0508	1.11
	60	1.3008	1.1472	0.3789	16.9396	1.2645	-0.2915	0.4213	17.9378	1.21
	90	1.3255	1.1309	0.4018	24.9280	1.1506	-0.1690	0.3984	23.3182	1.29
	120	1.1570	1.0246	0.3600	26.5568	1.1285	0.0952	0.3535	33.6866	1.13
	150	1.1853	1.1016	0.3969	34.7983	1.0719	-0.1299	0.3571	35.4937	1.02
500	**20**	1.1616	1.0356	0.3656	4.0423	1.3377	0.4002	0.4023	6.7281	0.95
	40	1.2901	1.0633	0.4017	10.5387	1.3058	0.0203	0.4292	12.9373	1.22
	60	1.3235	1.2154	0.3789	16.2760	1.1700	0.0050	0.3857	15.3782	1.16
	90	1.2444	1.2269	0.3873	21.7161	1.1326	0.2150	0.3695	23.8126	0.99
	120	1.1190	0.8867	0.3515	27.3362	1.1078	-0.0885	0.3841	29.2464	1.17
	150	1.1305	1.2054	0.3589	31.6906	1.0459	0.0527	0.3692	34.1625	1.14
1000	**20**	0.9148	0.3853	0.3396	3.4375	1.5795	0.0141	0.4675	9.0969	1.11
	40	1.2209	1.0638	0.3968	10.1406	1.2297	-0.0828	0.4096	11.7404	1.27
	60	1.1235	1.4114	0.3458	11.8648	1.2080	-0.0238	0.4159	17.3668	1.06
	90	0.9336	0.8526	0.3199	14.5700	1.1566	0.1833	0.3954	23.8303	1.13
	120	1.1972	1.0900	0.3785	27.3581	1.1430	0.1176	0.3905	31.6380	1.17
	150	1.0740	0.8868	0.3602	31.6370	1.1056	-0.0552	0.3820	37.3843	1.07
5000	**20**	0.9887	1.0775	0.3219	8.3219	1.0553	0.0487	0.3719	6.9611	0.97
	40	1.2061	1.1691	0.3747	9.0433	1.0705	0.3450	0.3633	9.5540	1.25
	60	1.2881	1.1324	0.3895	16.7549	1.1574	0.1319	0.4003	16.3038	1.11
	90	1.0367	1.2240	0.3329	16.5811	0.8578	-0.0972	0.3173	15.2790	1.19
	120	1.0323	0.9809	0.3437	22.5777	1.0251	0.0398	0.3659	26.7080	1.08
	150	0.9874	0.9627	0.3318	26.4567	0.9387	0.1802	0.3344	28.9885	0.96

In Table 1, obtained parameter estimations are shown under the assumption for the lower discrimination parameter and for low, neutral and high ability parameters.

According to Table1, while the number of item increase for each sample size, the discrimination parameter decreases. Besides, low number of item (*k*=20, 40) provides higher discrimination value whereas high number of item provides lower discrimination value in Figure 1.

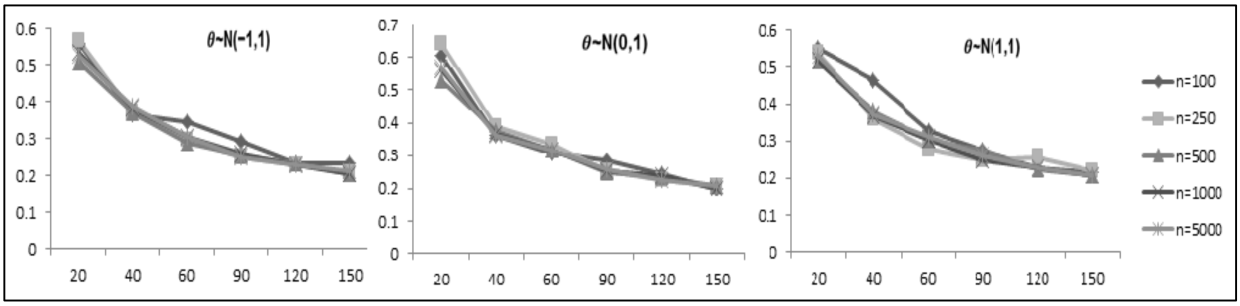

Figure 1. *The graphs of estimated discrimination parameters in different sample sizes and number of items for low, neutral and high ability levels.*

When the number of item increase for each sample size, point-biserial correlation coefficient decreases. Furthermore, the highest correlation coefficient values are obtained when the number of item is low. On the other hand, the lowest coefficient values are obtained when the number of item is high. There are not many differences among the point-biserial correlation coefficients when the sample size increases for each of the item numbers in Figure 2.

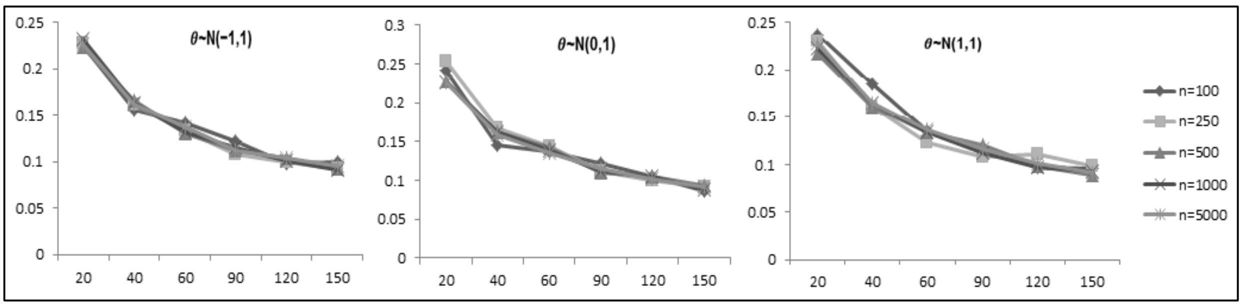

Figure 2. *The graphs of point- biserial correlation coefficient in different sample sizes and number of items for low, neutral and high ability levels.*

When the number of sample size increases, TIF value generally decreases. Moreover, TIF has the highest value when the number of item is the highest (*k*=150). A general increase or decrease can not be seen in the discrimination parameter when the number of item increases for each of the sample sizes. A general decrease is observed for all item numbers when the sample size increases as it is seen in Figure 3.

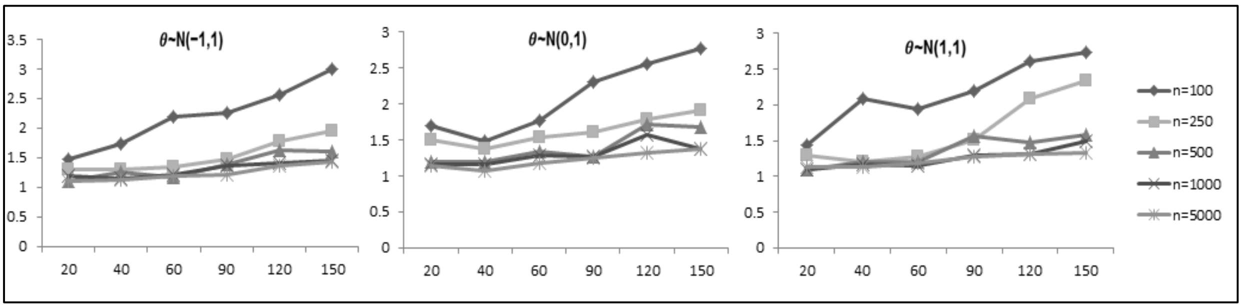

Figure 3. *The graphs of test information function in different sample sizes and number of items for low, neutral and high ability levels.*

In Table 2, obtained parameter estimations are shown under the assumption that the discrimination parameter of item is higher and ability parameters are low, neutral and high. According to Table 2, a general increase or decrease can not be seen in the discrimination parameter of item when number of item increases for each of the sample sizes.

A higher discrimination can be seen when sample size is less (*n*=100, 250). On the contrary, the discrimination is observed lower in the situation where sample size is higher. In general, low number of item (*k*=20, 40) provides a high discrimination as it is seen in Figure 4.

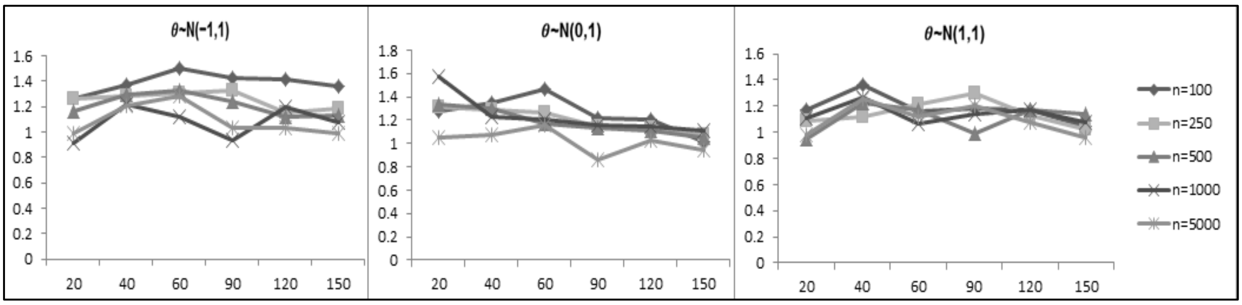

Figure 4. *The graphs of estimated discrimination parameter in different sample sizes and number of items for low, neutral and high ability levels.*

It is observed that point-biserial correlations are approximately equal for most of the sample sizes and the item numbers in terms of three ability levels.

TIF increases when the number of item increases for each sample size. In addition, high number of item produces the highest TIF. In general, TIF decreases for each item number when the sample size increases as it is seen in Figure 5.

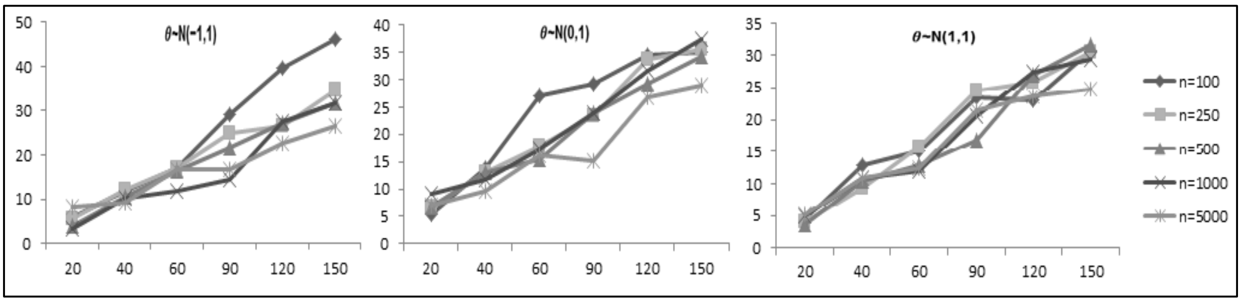

Figure 5. *The graphs of test information function in different sample sizes and number of items for low, neutral and high ability levels.*

6. NUMERICAL EXAMPLES

A hypothetical item response data is generated on the purpose of illustrating the utilization and applicability of the study for 20 item and 30 examinees. Using the full data set of 30 examinees and 20 item parameters from which 10 examinees and 10 item parameters were extracted and shown in Table 6.1. The data set was coded as 0=incorrect and 1=correct.

Table 6.1. *Hypotetical item response data*

Item	Examinee										Total Score
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
1	0	0	0	1	0	0	0	0	1	1	3
2	0	1	0	1	0	0	0	0	1	1	4
3	0	0	0	1	0	0	1	0	0	1	3
4	0	0	0	0	0	0	0	0	0	0	0
5	0	1	0	1	0	0	0	1	0	0	3
6	0	0	0	0	1	0	0	0	0	0	1
7	0	0	0	1	0	1	1	0	0	1	4
8	1	1	1	1	0	0	1	0	0	1	6
9	0	0	0	0	0	1	0	0	0	0	1
10	0	0	0	1	0	1	1	0	1	0	4

It is assumed that item discrimination (*a*) and difficulty (*b*) parameters were already known before starting parameter estimation. Initial parameter values were determined having regard to high discrimination parameter (*a*) and low ability level for the theta. The initial item discrimination and difficulty parameters values for 20 items are shown in Table 6.2.

Table 6.2. *Initial parameter values of item discrimination (a) and difficulty (b)*

k	*a*	*b*	*k*	*a*	*b*
1	0.5273	0.7378	11	1.4318	0.5454
2	0.4696	-0.3414	12	0.5615	0.6872
3	1.6793	0.8886	13	0.8245	0.3720
4	0.9911	-1.1797	14	0.7244	-1.4205
5	0.3047	1.3815	15	1.5628	-0.3357
6	0.4615	1.1604	16	0.2710	-1.9762
7	1.3159	0.0185	17	1.8041	-0.3830
8	1.1259	0.4039	18	0.5793	-0.7899
9	0.5837	0.9795	19	0.9991	-0.0960
10	1.2446	-0.3277	20	1.5672	0.3774

Table 6.3 shows that the initial examinee ability parameter values of 30 examinees.

Table 6.3. *Initial values of examinee ability parameters*

n	***theta***	***n***	***theta***	***n***	***theta***
1	0.3014	**11**	-0.8734	**21**	-0.5786
2	-0.9146	**12**	1.1150	**22**	0.4224
3	-0.5211	**13**	-1.9942	**23**	-0.3119
4	-1.2628	**14**	-2.5846	**24**	-0.0870
5	0.5868	**15**	-0.6892	**25**	-2.1189
6	-1.4040	**16**	-1.9696	**26**	-0.7592
7	0.1517	**17**	-1.5892	**27**	0.1234
8	0.1531	**18**	-0.6686	**28**	-1.6482
9	-1.5256	**19**	-2.2981	**29**	-0.9126
10	-0.5706	**20**	0.0717	**30**	0.6347

Furthermore, Table 6.4, Table 6.5, Table 6.6 and Table 6.7 show parameter estimation values for item discrimination (a) and difficulty (b) parameters; examinee ability (theta) parameter; point-biserial correlation (r_{pbi}) and Test Information Function (TIF).

As it is seen in Table 6.4, Item 8 has the highest discrimination value whereas Item 14 has the lowest value. Similarly, it is observed that Item 8 is the most difficult item while Item 16 is the easiest item.

Table 6.4. *Estimated parameter values of discrimination (a) and difficulty (b)*

k	***a***	***b***	***k***	***a***	***b***
1	0.9560	1.6320	**11**	1.7098	2.0756
2	0.7648	0.6244	**12**	0.5446	1.0779
3	1.3084	2.0860	**13**	1.0549	1.6810
4	2.4708	-0.9869	**14**	0.0719	-0.2686
5	0.1028	0.8494	**15**	2.1318	1.3127
6	0.0822	1.0131	**16**	3.9038	-4.2867
7	1.3242	1.3698	**17**	2.9530	0.8574
8	4.3161	4.0730	**18**	0.9955	-0.8316
9	1.6811	1.3267	**19**	1.2232	0.2054
10	0.9560	0.6636	**20**	1.2689	1.5615

As it is seen in Table 6.5, it can be inferred that the examinees having the highest ability are number 22 and 30. The examinees having the lowest ability are number 6,13,16,17 and 19.

Table 6.5. *Estimated parameter values of examinee ability (theta)*

n	***theta***	***n***	***theta***	***n***	***theta***
1	0.5519	**11**	0.5519	**21**	-0.1418
2	0.1165	**12**	1.1324	**22**	1.6256
3	0.3439	**13**	-1.5797	**23**	-0,8725
4	-0.4533	**14**	-0.8725	**24**	1.3560
5	0.9356	**15**	0.3439	**25**	-0.4533
6	-1.5797	**16**	-1.5797	**26**	0.5519
7	0.5519	**17**	-1.5797	**27**	0.5519
8	0.7467	**18**	0.1165	**28**	-0.1418
9	-0.8725	**19**	-1.5797	**29**	-0.8725
10	0.3439	**20**	1.1324	**30**	1.6256

It is seen in Table 6.6 that the correlation between Item 17 and the total scores is a high value (r_{pbi}=0.6900). The point-biserial correlation coefficient indicates that Item 17 discriminates examinees well in this group.

Table 6.6. *Estimated values of item point-biserial correlation (r_{pbi})*

k	$\boldsymbol{r_{pbi}}$	***k***	$\boldsymbol{r_{pbi}}$
1	0.2770	**11**	0.4680
2	0.3092	**12**	0.2333
3	0.3588	**13**	0.3247
4	0.6620	**14**	0.1002
5	0.1334	**15**	0.6281
6	-0.0259	**16**	0.2341
7	0.4061	**17**	0.6900
8	0.6590	**18**	0.4052
9	0.5086	**19**	0.4863
10	0.3687	**20**	0.4110

Table 6.7. *The estimation values for examinee Test Information Function (TIF)*

n	***TIF***	***n***	***TIF***	***n***	***TIF***
1	9.5040	**11**	9.5040	**21**	6.1158
2	7.3289	**12**	9.4536	**22**	4.9493
3	8.4019	**13**	1.4294	**23**	2.9992
4	4.7022	**14**	2.9992	**24**	7.1274
5	10.7430	**15**	8.4019	**25**	4.7022
6	14.294	**16**	1.4294	**26**	9.5040
7	9.5040	**17**	1.4294	**27**	9.5040
8	10.5362	**18**	7.3289	**28**	6.1158
9	2.9992	**19**	1.4294	**29**	2.9992
10	8.4019	**20**	9.4536	**30**	4.9493

TIF of examinee 5 has the highest value (10.7430) in Table 6.7. It can be inferred from the results that examinee 5 has the highest ability.

Table 6.8. *Mean of the parameter estimation values*

a	***b***	r_{pbi}	***TIF***
1.4902	0.5018	0.4819	6.1792

When Table 6.8 is investigated, it is seen that the item discrimination is high and the item difficulty is low in terms of mean of the parameter estimation values. Mean point-biserial correlation indicates a moderate degree of relationship, so each item is seperating better examinees on the whole test from the weaker examinee adequately.

7. CONCLUSION

In the study, the aim is to reveal the effect of different sample sizes, item numbers and ability levels on the estimation ability and difficulty parameters as well as test information function and point-biserial correlation for 2-PL IRT model. In addition, low and high item discrimination levels are studied separately to understand how item discrimination level is able to be efficient on estimation results.

The simulation results shows that the level of discrimination parameter plays an important role in the parameter estimation for 2-PL IRT models and three different examinee ability levels produce significant results. One of the significant results is the decrease in the item discrimination value when the number of item increases for all sample sizes in the predetermined low discrimination level. Besides, the item discrimination value is quite low when the number of item is high. On the other hand, the highest discrimination value is observed for the lowest number of item (k=20). In general, the discrimination of item is the highest for the examinees with low ability for all sample sizes and the number of items when the discrimination of item is the lowest for the examinees with high ability.

The highest difficulty is obtained for all sample sizes and the number of items when the ability level is low in predetermined high discrimination level. On the contrary, the lowest difficulty value is obtained when the ability level is high. In general, the difficulty value of item decreases when the number of item increases.

It is seen that TIF increases for the lowest sample size when the number of item increases in predetermined high discrimination level. On the contrary, TIF decreases for the low sample size (n=100, 250) when the number of item increases in predetermined low discrimination level.

On the other hand, point-biserial correlation increases for all sample size when number of item is lower in the predetermined low discrimination level. It is an expected situation that the obtained information will be much more in a study containing less examinees. In addition, high number of item can be helpful to obtain more accurate and detailed information.

In application study, determination of the item discrimination value becomes difficult when the number of item increases. Point-biserial correlation coefficient is important in determination whether item discrimination value is low or high. Since high item number causes decrease in point-biserial correlation, low item number is a suggestion to the researchers. Tests with 40 items reveal good estimation in this study.

The item discrimination parameter has an importance role to compute item information function. Test information function value contributes a significant benefit to the researchers while making decision on which examinee has higher ability level. It can be inferred from the results that high discrimination parameter value provides high item information function whereas low dicrimination value causes a low item information function. Briefly, item discrimination and difficulty parameters, examinee ability level, point-biserial correlation and test information function are considerably important for the researches in all fields.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors

REFERENCES

[1] Andersen, E. B., “Sufficient Statistics and Latent Trait Models”, Psychometrica, 42, 69-81, (1977).

[2] Baker, F. B., The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, (2001).

[3] Baker, F. B. and Kim, S. H., Item Response Theory: Parameter Estimation Techniques, Second Edition, (2004).

[4] Baur, T. and Lukes, D., 2009. An Evaluation of the IRT Models Through Monte Carlo Simulation, Journal of Undergraduate Research XII.

[5] Birnbaum, A.,Some Latent Trait Models and Their Use in Inferring an Examinee’s Ability, In F.M. Lord & M.R. Novick (Eds.), Statistical Theories of Mental Test Scores, (1968).

[6] Cai, L., and Thissen, D., Modern Approaches to Parameter Estimation in Item Response Theory. In S. P. Reise & D. A. Revicki (Eds.), Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment (pp. 41-59). New York, NY: Routledge, (2014).

[7] Hambleton, R.K.,“Item Response Theory: Introduction and Bibliography”, Psicothema, 2(1), 97-107, (1990).

[8] Harris, D., Comparison of 1-,2- and 3-Parameter IRT Models, Instructional Topics in Educational Measurement, An NCME Instructional Module on, (1989).

[9] Hulin, C. L., Drasgow, F., and Parsons, C. K., Item Response Theory: Application to Psychological Measurement, Homewood, I11: Dow Jones-Irwin, (1983).

[10] Mellenberg, G.J., “Generalized Linear Item Response Theory”, American Psychological Association, 115 (2), 300-307,(1994).

[11] Le, D. T., Applying Item Response Theory Modeling in Educational Research, Iowa State University Digital Repository, (2013).

[12] Paolino, J. P., “Penalized Joint Maximum Likelihood Estimation Applied to Two Parameter Logistic Item Response Models”, Graduate School of Arts and Sciences, Columbia University, (2013).

[13] Rasch, G., Probabilistic Models for Some Intelligence and Attainment Tests, Chicago: MESA, (1960).

[14] Rizopoulos, D., “ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses”, Journal of Statistical Sofware, 17 (5), (2006).

[15] Toribio, S. G., “Bayesian Model Checking Strategies for Dichotomous Item Response Theory Models”, Graduate College of Bowling Green State University, (2006).