Accuracy of Bayesian method estimating for a person's ability and item parameters - two-parameter logistical model (2-PL) of different sample sizes

Shadi Yousef Shawawreh1[1*]
[1]Faculty of Art and Sciences, Aqaba University of Technology, Jordan
*shadi9003@yahoo.com

ABSTRACT

This study aimed to show the accuracy of estimating the Bayesian method in estimating item parameters and person's ability parameter under any circumstances relating to the sample size, Four levels of the sample were used: 1500,500, 200,100 by making a comparison between the Bayesian method of estimation that is used in the software (WinBUGS V1.4) and the maximum likelihood-is the usual to estimate parameters -way by (Bilog-MG3) software according to a two-parameter logistic (2-PL) model in the item response theory (IRT) models -dichotomous response, Experimental data were generated using WinGen v.3 software, and the item parameters and ability were estimated according to the two methods(Bayesian, maximum likelihood). The results of the estimation accuracy of the Bayesian method were reached in light of the circumstances use of the sample size variable by evaluating the Root Mean Squared Error (RMSE) index of the arithmetic mean, and the relative efficiency (RE) index and the results showed the advantage of the Bayesian method of estimation in small samples, and this was confirmed by the results of (t-test) to compare the correlation coefficients at the two methods of estimation between the generated values and estimated values of the parameters for each method, where parameters showed The correlation of the Bayesian method is a statistically significant difference at (α= 0.05).

Keywords
Bayesian method, maximum likelihood, two-parameter logistic

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

Introduction and theoretical framework

The study of the inherent features of individuals and their measurement by the Classical Test Theory (CCT) began in the 1920s, which showed some deficiencies in its procedures; Therefore, the Item Response Test theory, (IRT) has emerged, and among its assumptions is the ability to predict the performance of individuals or explain their performance through a test that focuses on the distinct characteristics of this performance. These are called inherent features or abilities, which cannot be observed directly but rather from the performance of individuals through a group of items. Therefore, the item response theory aims to determine the probable relationship between an individual's performance in a test and the underlying feature of this performance. This relationship is expressed in logistical models, and these models attempt to estimate the position of the individual (examined) on the trait of the attribute or ability using the pattern of its responses. The theory (IRT) is based on an underlying assumption that the probability of the individual responding to any of the test items correctly, is a combination of both the individual ability that the test measures on the one hand and the characteristics of the paragraph that the individual is trying to answer on the other hand, and this requires obtaining information about the individual and the item, I.e. obtaining estimates of the capabilities of individuals and the features of the items (Crocker & Algina, 1986).

IRT models and equations differ according to the parameters of the items to be estimated. The one-parameter logistic model proposed by George Rasch (1960) in turn estimates the parameter of difficulty and individual capacity, and the Birnbaum Model proposed by (Birnbaum, A.1968) estimates the parameters of the items (difficulty and discrimination) and the ability of the individual, while the three-parameter logistical model proposed by Pernium in 1968 AD after adding the guesswork parameter to the binary model, estimates the parameters of the items (difficulty, discrimination and guesswork) and the ability of the individual. These models are the most common models of Item Response Theory and are most appropriate for Dichotomous Models. In this study, the Two-Parameter Logistic Model (2-PL) was used. It is represented by the following mathematical function:

$$p_i(\theta) = \frac{1}{1+e^{-ai(\theta_j - b_i)}}$$

Since pi (θ) is the probability of the subject's answer, which is his ability on paragraph i, which is difficult as (i =1,2, ……, n) and n is the number of items in the test (Baker, 2001).

Methods for estimating the parameters of items

Estimating the parameters of the items is considered a fundamental pillar through which we can estimate the ability of individuals according to the theory (IRT), and given that the parameters of the items are unknown, we estimate them using one of the estimation methods, by relying on individuals' answers to the test items, and the process of determining the parameters of the items and the ability in IRT theory models needs accuracy in estimating with the least possible estimation error, this depends on the procedures used in estimating the parameters of the item and ability. Since there are several methods for estimating according to IRT theory, Hambleton & Swaminathan (1985) pointed out the following methods:

Bayesian Modal Estimation

The Bayesian method in statistical inference belongs to the British scientist Thomas Bayes (Bellhouse, 2004). This is one of the methods used to estimate the parameters of the items in IRT theory. Using this method, the parameter to be estimated will be considered as a random variable that follows a specific probability distribution, or what is known as Prior Distribution, by taking into account the previously unknown information of the parameter during the estimation process. The previous information is usually determined by the researcher relying on a personal belief That is, based on his previous experience, or through the statistical characteristics of that parameter which will be estimated (Lord, 1986) (Swaminathan & Gifford, 1982).

Gao & Chen (2005) pointed out that Bayesian estimates are based on Prior distribution or what is called primary information. If prior distribution lacks the information, then Bayesian estimates and the Maximum Likelihood have comparable results. However, if the distribution is a normal distribution, the Bayesian method is more accurate. There are those who believe that the Bayesian method is more accurate in estimating compared to other methods such as MLE, Finch & French, 2019. Swaminathan & Gifford (1982) as well as in the triple logistic model (Swaminathan & Gifford 1986 (Hambleton & Swaminathan, 1985), Bayes' equation is presented as follows:

$$f(\theta,a,b,c|u)\infty L(u|\theta,a,b,c)\left\{\prod_{i=1}^{n} f(a_i)f(b_i)f(c_i)\right\}\prod_{a=1}^{N} f(\theta_a) \quad ....(24)$$

($f(\theta_i), f(a_i), f(b_i), f(c_i)$) indicates the dimensional distributions of the features, as:

The ability of individuals (aθ) distinguishing of the item (ai) The difficulty of the item (bi) item guessing (ci), and the parameter can be added or deleted based on the logistical model used, and it is required to specify the power distributions (θ) to find the parameters of item using the Bayesian method, and estimates of parameters values (θ, a, b, c) that maximize typical posterior distribution are found, thereby obtaining an estimate of the parameters of the items.

Maximum Likelihood Estimation (MLE)

It is one of the leading and most prevalent methods for estimating the parameters of the item and ability, and this is done through the procedures for maximizing the probability of the parameter to be estimated. This method is used in (Bilog. Mg- 3) program. (Rijmen, 2009) (Hambleton & Swaminathan, 1985) has demonstrated that this method is based on finding the estimation of parameters through the probability-maximization procedures for the ability parameter to be estimated when we have sample information. The maximum likelihood method for estimation has several types that differ according to the method used in the estimation process, including Joint Maximum Likelihood (JML), Conditional Maximum Likelihood (CML), and lastly, Marginal Maximum Likelihood (MML) which will be adopted in this study, because it is approved in the (Bilog-MG 3) program, where the ability values of individuals will be estimated according to this program.

MML method for estimation was developed by Bock and Aitkin (1981). This method was used to estimate the parameters of items and ability in the program (Bilog-MG 3). The thing that distinguishes the MML method is that it addresses some of the issues found in the previous two methods (JML and CML), Where it estimates the parameters of the items and the ability of the mono, bilateral and triple logistical models (1- PL, 2-PL, 3-PL) and deal effectively with the number of items, whether they are few or many, and this is where the CML method failed to do. In addition, the MML method deals with all test items, whether they have been answered by individuals or not. It also addresses the problem of instability in the JML method, which results from estimating the parameters of individuals and items together. The most distinguishing feature of this method is that it gives the least estimation error of ability compared to other likelihood methods. (Hambleton & Swaminathan, 1985) (Lo, Liu & sheo, 2003). The (MLE) method for dealing with small samples according to the mono logistic model (1-PL) gives results which are less accurate than the Bayesian method due to the increase in the error rate of estimation (Azizan, Mahmud, Rambli, 2019 (and it successfully handled small and abnormal sample data issues and produced a more accurate parameter estimate with the smallest squared error (MSE), especially in a small sample compared to MLE.

Hambleton & Swaminathan, 1985 have pointed that the procedures through which the estimation of items parameters and capacity in accordance with (MML) method pass two phases: the stage of expectation (Expectation Stage), and the stage of deification (Maximization Stage), and each stage has its own calculations to estimate. Baker and Kim, 2000 pointed to the mathematical function of maximum likelihood for ability estimation as follows:

Probability maximization equation (log-likelihood (LL) function)

$$LL(\theta,\xi)=\sum_{i=1}^{n} u_{ij}log[p_{ij}(\theta)]+(1-u_{ij})log[Q_{ij}(\theta)]$$

uj = response pattern for person j,
ξ = the item parameters for the administered item(s),
n = the number of items that have been administered,
Pij = probability of a keyed response,
$Qij = 1 - Pij$, or the probability of a non-keyed response, and
uij = item response

A procedure such as Newton-Raphson should be used to determine the maximum probability maximization, and iterative (Embretson & Reise, 2000). After that, the ability is estimated (estimate of θ)

$$\hat{\theta}_{i+1}=\hat{\theta}-\frac{\frac{\partial(LL)}{\partial(\theta)}}{\frac{\partial^2(LL)}{\partial^2(\theta)}}$$

Problem Statement

Accuracy of measurement is one of the most important objectives of the research field in measuring and evaluating the preparation of tests and metrics. The classic theory (CCT) in measurement was concerned with the accuracy of measurement through what are called psychometric properties (honesty and stability factors), and with the emergence of modern theory in measurement (IRT) there was a goal to reduce the margin of error in the measurement in order to increase accuracy and to determine the factors affecting the accuracy of the estimation of parameters of the item and the capabilities of individuals (discrimination (a), difficulty (b), estimation (c) and the ability of individuals (θ)). Opinions differed regarding the factors affecting the accuracy of the estimation of parameters, including the method of estimation and the number of individuals in the sample. Therefore, this study specifically seeks to determine the accuracy of the evaluation of the Bayesian method in estimating the parameters of the items and ability depending on the sample size according to the binary logistic model.

Swaminathan & Gifford (1982) conducted a study to compare the Bayesian method and the Maximum Likelihood method for estimating the parameters of items and ability according to the mono logistic model, using data generated by the DATGEN program for two samples of applicants (n = 50,75) and a test that was divided into three lengths as follows (n = 50,100,150). The study concluded that there are differences between the two estimation methods, especially when sampling sizes and the number of items is small as Bayes' estimates showed improvement in estimating the parameters of the items when the sample sizes were small.

Lord (1986) conducted a study to explain the method of Maximum Likelihood and Bayes in estimating the parameters of items and pointed out that Bayes' estimates based on previous results, reduce the mean square error (MSE) of the estimate compared to the maximum weight value due to our prior knowledge of item parameters. Based on this, we can determine the importance of the Bayes method for realistic tests, where previous distributions of tests that have been applied repeatedly on individuals can be obtained, as it becomes possible to deduce previous distributions of the parameters of the item and ability from previous results, and in this case, the Bayes procedures can give better estimates than the Maximum Likelihood method.

Item Parameter Estimates for Generated Data

The data was generated by the program (WinGen v.3) according to the two-parameter logistic model (2-PL). Table 1 shows The mean and the standard deviation of the values based on one of the two-person Item models at each level of the sample size variable (1500, 500,200,100).

Glass, 2005 conducted a study aimed to determine the effect of the sample size (500, 1000, 2000) and the number of items (440,200) on the accuracy of the estimation of the individual capacity parameter according to the Bayes method. Moreover, in order to achieve the objective of the study, two-steps data were generated according to the sample size and the number of test items and ability levels, and the results showed that the increase in the number of items at the sample size (2000,500) reduces the standard errors in the estimate, which increases the accuracy of the estimate, while the accuracy of the estimate decreased when the number of items at the size of a sample (1000) increased.

Gau & Chen (2005) also conducted another experiment aimed at comparing the MML method and Bayes method in estimating the parameters of the items according to the sample size (2000,500,100) and the number of items (60,30,10) according to the triple logistic model. The results showed a preference for the Bayesian estimation method when the sample size was (100), and the values were similar in terms of accuracy between the two estimation methods in different situations according to the RMSD equation.

Burgos (2010) conducted a study aimed at introducing the Bayes method in the field of item response theory (IRT). The researcher applied his study to the single parameter logistic model (Rush model) through simulation data with a sample size of (500) and a test with a number of items (11). The results concluded that the Bayes method gives more appropriate and convincing estimates of the parameters of the items and individuals in the (IRT) models than the methods of (MML, CML, JML). The ability of the Bayes method is less or equal to the maximum likelihood method, which gives preference to the Bayes method in estimating the parameters of the item.

Karadavut (2017) conducted a study aimed at comparing the effect of the ability type distribution (normal and uniform distribution) on estimating the ability according to the mono, bilateral, and triple logistic model (1-PL, 2-PL, 3-PL), with the different item number (15,30) and a sample size of (600,2000). The study found that Pearson correlation values and MSE and RMSE indicators specified that the uniform distribution was more accurate for the ability estimates in (2-PL,3-PL) logistical models, while there are no apparent differences between the effect of the standard and uniform distribution on ability estimates in (1- PL) logistic model.

Table 1. Descriptive Statistics for Item Parameters and (Real) Generated Capacity for a 50-item test according to the binary logistic model

Model	Number of Items	Parameter	Mean	Standard Deviation

Binary logistic model-2PL	50	Parameter of true preference	1. 113	0. 189
		Parameter of true difficulty	0. 217	1. 042
		True ability parameter of individuals	0. 214	1. 039

Estimate the parameters of the item and the ability of individuals using BILOG-MG3, WinBUGS v 1.4

The WinBUGS v 1.4 software was adopted to estimate the parameters of the items and the ability parameter for individuals in the Bayesian method. The data was retrieved using WinGen v.3 software which was developed by Han (Han, Hambelton, 2007), and then the data was passed to the software (WinBUGS v1.4) (Spiegelhalter, Thomas, & Best. 2003). To learn more about the software, Cowles (2004) is a good reference. Item parameters and ability parameter for individuals were estimated using Bilog-MG 3 software (MLE). To learn more about Bilog-MG 3 software, Toit (2003) is a good reference to the Operating Manual, Learn more about BILOG-MG3 software (Zimowski, Muraki, Mislevy, & Bock. 1996).

Statistical Analysis

The data for this study was generated using (WinGen-3) program, to conduct analyzes using (WinBUGS V 1.4), Bilog-MG3 and SPSS software. Moreover, to answer the research questions, the following statistical treatments were performed:

1. Calculating the mean and standard deviation for the parameters of items and the ability of individuals using the two estimation methods (MLE and Bayesian).

2. Calculating the square root for error rate square (RMSE), which is an indication of the estimation accuracy, where the index mean for the parameters of item parameters and the ability of individuals for each estimation method (MLE, Bayesian) was calculated.

$$RMSE_x = \sqrt{\frac{\sum_{i=1}^{n} (\hat{X}_{estmated} - X_{true})^2}{n}}$$

3. Calculating the relative efficiency index (RE), which is the variance of the estimated values, parameters of items, and the ability of individuals according to the two methods of estimation (MLE and Bayesian) in different situations of the sample size variable, using the following formula:

$$RE_x = \frac{V\ \hat{x}\ \text{Bayesian}}{V\ \hat{x}_{MLE}}$$

Whereas, the symbol "V" means the variation of the estimates of parameters according to the chosen estimation method.

4. Pearson correlation coefficient: The correlation coefficient was used to determine the strength of the relationship between the estimated values for each method of estimation with the values generated (real).

5. (T-test) test for correlation coefficients of non-independent samples: with the aim of knowing the statistical significance of correlation coefficients between estimated and real values according to each of the two estimation methods.

The arithmetic mean and standard deviations were calculated for the estimations of the parameters of the items and the ability of individuals and their estimated standard errors according to the two estimation methods (MLE and Bayesian) depending on the variable of the sample size, as shown in Table (2).

Table 2. Arithmetic mean for estimating Item and Ability Parameters and Standard Estimated Errors according to the two methods (MLE and Bayesian) based on the sample size variable

Model	Sample Size	Statistic	Estimation Method: Bayesian Mean	Bayesian Standard Deviation	MLE Mean	MLE Standard Deviation
Binary parameter	100	Preference parameter	0.868	0.45	1.022	0.61
		Standard error of preference	0.14	0.05	0.24	0.17
		Difficulty parameter	-0.465	0.89	-0.396	0.71
		Standard error of difficulty	0.25	0.11	0.25	0.06
		Ability parameter	0.000	0.79	-0.729	0.78
		Standard error of ability	0.10	0.02	0.08	0.03
	200	Preference parameter	0.738	0.25	0.799	0.26
		Standard error of preference	0.13	0.03	0.17	0.05
		Difficulty parameter	-0.423	0.92	-0.388	0.80
		Standard error of difficulty	0.18	0.07	0.19	0.08
		Ability parameter	-0.190	0.67	0.008	0.61
		Standard error of ability	0.17	0.02	0.15	0.03
	500	Preference parameter	0.782	0.29	0.839	0.31
		Standard error of preference	0.07	0.03	0.10	0.05
		Difficulty parameter	-0.479	1.22	-0.501	0.96
		Standard error of difficulty	0.17	0.13	0.13	0.09
		Ability parameter	0.012	0.82	-0.529	0.79
		Standard error of ability	0.22	0.01	0.13	0.02
	1500	Preference parameter	0.743	0.21	0.780	0.22
		Standard error of preference	0.08	0.05	0.11	0.04
		Difficulty parameter	-0.419	1.18	-0.442	0.95
		Standard error of difficulty	0.11	0.13	0.15	0.07
		Ability parameter	-0.711	0.75	0.030	0.81
		Standard error of ability	0.09	0.03	0.06	0.04

To find out the accuracy of the estimate, the arithmetic mean for the RMSE indicator was calculated as an indicator of estimation accuracy for the parameters of items and the ability of individuals for the binary model (2-pl) according to the different estimation method (MLE, Bayesian), based on the sample size variable, as shown in Table (3).

Table 3. arithmetic mean for the RMSE index in the accuracy of estimating the preference, honesty and ability parameters according to the different estimation method (MLE, Bayesian), based on the sample size variable.

	Sample Size

Paramete r	RMSE Accuracy Estimation for (Distinction, Difficulty, and Ability) parameters	100	200	500	1500
Preferenc e	Bayesian	0. 387	0. 352	0. 378	0. 403
	MLE	0. 465	0. 393	0. 385	0. 363
Honesty	Bayesian	0. 319	0. 325	0. 259	0. 343
	MLE	0. 328	0. 339	0. 242	0. 317
Ability	Bayesian	0. 312	0. 348	0. 311	0. 309
	MLE	0. 373	0. 325	0. 283	0. 298

It is noted from Table (4) that all RMSE values as a precision indicator of Bayesian method for parameter estimates based on sample size (100, 200, 500) were apparently smaller than the MLE estimate method; This means that the Bayesian method was ostensibly more accurate in estimating the preference parameter in most sample sizes. Similarly, the RMSE index as an indication of the accuracy estimate of the difficulty parameter according to the Bayesian estimation method based on sample size (100, 200) was apparently smaller than the MLE estimation method. While the RMSE index, as an indication of the accuracy estimate of the ability parameter according to MLE and Bayesian estimation methods, was ostensibly more accurate in estimating the ability parameter in sample size (100, 500). Furthermore, MLE’s advantage in estimating the power parameter at a sample size of (200). The researcher attributed this result to the coincidence of the unreliability of the MLE estimates for small samples and was referred to in the operating manual for the Bilog MG-3 program (DU, Toil, M. 2003). It should be noted that MLE estimates are not reliable for small samples especially when the sample size is 250 individuals and less, and this was mentioned in the operating manual for Bilog MG-3 (DU, Toil, M. 2003).

The variance of the estimated values of the preference parameter, difficulty and ability of individuals according to different estimation methods (MLE and Bayesian) was calculated with different sample size numbers using the relative efficiency index according to the mathematical formula

Table 4. values of the relative efficiency of the preference parameter, difficulty and ability of individuals according to the two methods of estimation (MLE and Bayesian) based on sample size variable.

Parameter	Statistic	Sample Size			
		100	200	500	1500
Preference	Preference parameter variation by Bayesian method	0.211	0.065	0.082	0.073
	Preference parameter variation by MLE method	0.382	0.072	0.150	0.068
	The relative efficiency of Bayesian on MLE	**0.552**	**0.090**	**0.547**	**1.073**
Difficulty	Difficulty parameter variation by Bayesian method	0.497	0.616	0.062	0.088
	Difficulty parameter variation by MLE method	0.764	0.839	0.051	0.070
	The relative efficiency of Bayesian on MLE	**0.651**	**0.734**	**1.216**	**1.257**
Ability	Ability parameter variation by Bayesian method	0.501	0.721	0.059	0.091
	Ability parameter variation by MLE method	0.767	0.711	0.054	0.079
	The relative efficiency of Bayesian on MLE	**0.653**	**1.014**	**1.093**	**1.151**

It is noted from Table (4) that the relative efficiency values of the parameters of the items (Preference, difficulty and ability) according to the Bayesian method on MLE, were apparently in favour of the Bayesian estimation method when the sample size is small (100, 200), except in a situation when the relative efficiency of a parameter was found The ability at a sample size (200) was apparently in favour of MLE, this result coincidentally due to the unreliability of MLE estimates for small samples as previously mentioned, while it is noted from Table (4) that the relative efficiency values of the Bayesian method on MLE for the item parameters (preference and difficulty) The capacity was in favour of the MLE method when the sample size was large (500, 1500).

To find out more about either of the two estimation methods (Bayesian MLE), we calculated the values of the correlation coefficients between the parameters of the items (preference and difficulty) and the actual ability, and the parameters of the items (preference and difficulty), and the estimated ability of the binary model according to the two estimation methods (Bayesian and MLE) based on the difference of sample size, then t-test was used for correlation coefficients of correlated samples to reveal the essence of the difference in the values of (correlation coefficient according to the MLE method) and (correlation coefficient according to the Bayesian estimation method) in each interactive position of the sample size (As shown in Table 5).

Table 5. Results of the t-test for the correlation coefficients between the preference, difficulty, and ability parameter according to the two estimation methods based on the sample size

Parameter	Sample size	Correlation between the true preference parameter and the (preference, difficulty and ability) parameter estimated according to the estimation method:		Correlation of the parameter (preference, difficulty and ability) according to the two methods	Calculated value of t	Statistical significance
		Bayesian	MLE			
Preference	100	**0. 7954**	**0. 7834**	**0. 98**	**0. 479**	**0. 026**
	200	**0. 8501**	**0. 8369**	**0. 99**	**-4. 990**	**0. 000**
	500	**0. 8826**	**0. 8415**	**0. 99**	**-24. 997**	**0. 000**
	1500	**0. 7519**	**0. 7683**	**0. 98**	**-15. 043**	**0. 000**
Difficulty	100	**0. 9495**	**0. 9386**	**0. 99**	**-3. 500**	**0. 001**
	200	**0. 9625**	**0. 9617**	**1. 00**	**-0. 941**	**0. 047**
	500	**0. 6627**	**0. 6918**	**0. 99**	**7. 596**	**0. 000**
	1500	**0. 5439**	**0. 5572**	**1. 00**	**9. 125**	**0. 000**
Ability	100	**0. 9688**	**0. 9687**	**1. 00**	**-0. 346**	**0.017**
	200	**0. 9823**	**0. 9822**	**1. 00**	**-6. 541**	**0. 042**
	500	**0. 9745**	**0. 9746**	**1. 00**	**-5. 382**	**0. 120**
	1500	**0. 9810**	**0. 9812**	**1. 00**	**-7. 535**	**0. 000**

Table (5) shows that there was a statistically significant difference at the level of significance (α = 0.05) when the sample size was small (100,200,500) between the values of (correlation coefficient between the actual preference parameter and the preference parameter estimated according to the Bayesian estimation method) and (correlation coefficient between the actual preference parameter and the preference parameter estimated according to the MLE estimation method) in favour of the Bayesian estimation method; This may be attributed to the effect of taking into account previous distributions of parameter parameters in the probability estimates according to Bayesian method, especially when estimating the parameters of the items and the ability for small sample sizes, this result is consistent with what was indicated by (LORD, 1986), (Swaminathan & Gifford 1986). As for the sample size (1500), preference was given to the MLE method. On the other hand, there is a statistically significant difference at the significance level (0.05 = α) when the sample size was (100,200), results are in favour of (correlation coefficient between the real difficulty parameter and the estimated difficulty parameter according to the Bayesian estimation method). When the sample size was (500, 1500), the results were in favour of (correlation coefficient between the real difficulty parameter and the estimated difficulty parameter according to the MLE method). As for the ability values of individuals, the results showed that there is a statistically significant difference at the level of significance (α = 0.05) between the values of (correlation coefficient between the real ability parameter and the estimated ability parameter according to the MLE method) and (the correlation coefficient between the real ability parameter and the estimated ability parameter according to the Bayesian estimation method) when the sample size was (100, 200), in favour of Bayesian method while results indicated that there was a statistically significant difference at the level of significance (α = 0.05) when the sample size was (1500) individuals in favor of the MLE method, while in Table (5) there was no statistically significant difference at the level of significance (α = 0.05) Between the values of the correlation coefficient between the actual ability parameter and the estimated ability parameter according to the two estimation methods (Bayesian MLE) at a sample size of (500).

References

[1] Azizan, N, H, Mahmud, Z, Rambli ,A,.(2019). Rash measurement model: a review of Bayesian estimation for estimating the person and item parameters. Journal of Physics: Conference Series 1366 (1), 012105

[2] Baker, F. B., & Kim, S. H. (2004). Item response theory: Parameter estimation techniques. New York, NY: Marcel Dekker.

[3] Bellhouse, D.R. (2004). The Reverend Tomes Bayes, FRS: A Biography to

celebrate the Tercentenary of His Birth. Statistical Science 19 (1), 3- 43.

[4] Birnbaum, A. (1968). 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 (pp. 397-479).

[5] Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46(4), 443–459.

[6] Burgos, J. G. (2010). Bayesian Method's in Psychological Research. The case of IRT. International Journal of Psychological Research, 3(1), 164- 176.

[7] Cowles, K. (2004). "Review of WinBuGs 1.4" The American statistician, 58, 330-336.

[8] Crocker, L. and Algina, J. 1986. Introduction to classical and modern test theory, New York: Holt, Rinehart and Winston.

[9] Du, Toit, M. (2003). IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TEST-FACT. Lincolnwood, IL: Scientific Software International.

[10] Embretson, S. E. & Reise, S. P. (2000). Item response theory for psychologists. Mahwah,NJ: Lawrence Erlbaum Associates.

[11] Finch H and French. B F 2019 A Comparison of Estimation Techniques for IRT Models With Small Samples. Applied Measurement in Education 32(2):77–96

[12] Gao, F., & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. Applied Measurement in Education, 18, 351-380.

[13] Glass, G .(2005). The impact of item parameter estimation of computerized adaptive testing with item cloning. Law School Admission Council Computerized Testing Report 02-06 November.

[14] Hambleton, R.K.,& Swaminathan, H. (1985). Item Response Theory: Principles and Applications, Boston, Kluwer, Nijhoff Publishing.

[15] Han, K.T., & Hambelton, R.K. (2007).User Manual For WinGen: Windons Software That Generates IRT Model Parameters And Item Response. Holt Rinehart and Weston, New York.

[16] Karadavut,T.(2017). Estimation of item response theory models when ability is uniformly distributed. The Eurasia Proceedings of Educational & Social Sciences,7,30-37

[17] Lo, S., Liu, X. & Shao,Y. (2003).A Marginal likelihood for Family Bata . Annals of Human Genetics, 63,357-366.

[18] Lord, M. F. (1986). Maximum Likelihood, and Bayesian parameter Estimation in Item Response Theory. Journal of Educational Measurement, 23 (2),157-162.

[19] Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielson and Lydiche (for Dan marks Paedagogiske Institute).

[20] Rasch, G. (1960). Probabilistic models for some intelligence tests and attainment tests. Danish Institute for Educational Research, Copenhagen, Denmark.

[21] Rijmen, F. (2009). Efficient full in Formation Maximum Likelihood Estimation for Multidimensional IRT Models, Educational Testing service (ETS).

[22] Spiegelhalter, D., Thomas, A., & Best, N. (2003). WinBUGS (version 1.4) [Computer software]. Cambridge, UK: Biostatistics Unit, Institute of Public Health.

[23] Swaminathan H and Gifford J, A. (1986). Bayesian Estimation in the Three-Parameter Logistic Model. Psychometrika. 51(4).

[24] Swaminathan, H., & Gifford, J. , A. (1982). Bayesian in the Rasch Model. Journal of Educational Statistics, 7 (3), 175- 191.

[25] Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG: Multi-group IRT analysis and test

maintenance for binary items. Chicago, IL: Scientific Software International.