Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 8, 1744-1759

Accuracy Of Estimating Item Parameters And Individual' Ability Of Three-Parameter Item Response Theory Model Using Joint Maximum Likelihood Estimation Method In Light Of Sample Size And Test Length

Randa "Ahmad Fathi" Mohammad Al_alem

Assistant Professor In Education /Teacher Qualification Program At Palestine Technical University Kadoorie

Abstract

The current study aimed to reveal the accuracy of estimating the parameters of the items and the ability of Individual of the three-parameter item response theory **(**3PL**)** model using Joint Maximum Likelihood Estimation (JMLE) method according to the sample size and test length. To achieve this, Response data (100, 250, 500, 1000) were generated on tests consisting of (20, 40, 60) items, and the R programming language package (R-4.1.0) was used to estimate item parameters (difficulty, discrimination, guessing). For Individual' abilities, the JMLE method was used. The correlation coefficient indicator demonstrated that as the sample size and length of the test increased, the value of the correlation coefficient between the expected value and the estimated value of the item's parameters increased. The findings also showed that by increasing the sample size and length of the test, the value of the root mean square error index (RMSEA) fell, indicating that the estimated value of the item parameters and the person's ability was reduced. The study suggested comparing the accuracy of calculating item and person ability using the marginal maximum likelihood approach and the joint maximum likelihood method.

Keywords: Joint maximum likelihood method, three-parameter item response theory model, test length, sample size.

Introduction

Item Response Theory (IRT) contributed to finding many new mathematical models that were used in building and developing many psychological and educational measurement tools (Ogunsakin & Shogbesan, 2018; Dohoo & Emanuelson, 2021). IRT models can also be categorized based on the number of scored responses as dichotomous models and polytomous models. They were also classified according to the number of measured attribute dimensions to multidimensionality models, and unidimensional models, which are considered the most widely used by educational psychologists in building and developing educational and psychological tests and standards (Hambleton & Swaminathan, 1985). Dichotomous IRT models are described by the number of parameters they make use of the such as the one-parameter logistic model(1PLM), two-parameter logistic model (2PLM), three-parameter logistic model (3PLM), and four-parameter logistic model (4PLM), which represents the broader model of the previous models, and its mathematical equation that determines the relationship between an person's performance on an item and the ability are explained as follows (Baker and Kim, 2004; Osterlind & Wang, 2017; Philip & Ojo, 2017; Ogunsakin & Shogbesan, 2018):

$$\mathbf{P_i(\theta) = c_i + (d_i - c_i)\frac{1}{1 + e^{-1.702a_i(\theta - b_i)}}}$$

It is noted that it includes five parameters: the symbol of person ability is (θ), the item recognition parameter (a_i), the item difficulty parameter (b_i), the item guessing parameter is (c_i), and the item upper asymptote "carelessness" parameter (d_i).

When the value of item discrimination is one and its guessing is zero, and the value of item upper asymptote "carelessness" is one, then this model is called the one-parameter item response theory model (1PLM) or Rasch model, When the value of item guessing is zero, and the value of the item upper asymptote "carelessness" is one, then this model is called the two-parameter item response theory model (2PLM), When the value of item upper asymptote "carelessness" is one, then this model is called the three-parameter

item response theory model (3PLM), the four-parameter logistic model (4PLM) assumes that even high ability examinees can make mistakes (e.g. due to carelessness)(d_i < 1) (Świst, 2015; Ogunsakin & Shogbesan, 2018).

IRT is based on strong assumptions (Sijtsma & Junker, 2006; Liu et al., 2021; Temel et al., 2022; Mutiawani et al., 2022), namely; Unidimensional models that require a single trait (ability or simple structure) (Bulut, 2013; Heene et al., 2016), and local independence, which relates to the fact that person's response to one item is not affected by a response to another item (Kim et al., 2021). The existence of an Item Characteristic Curve (ICC) that describes the relationship between an person's correct answer to the item and the latent ability measured by the test or scale, and the mathematical form of ICC is the logistic form whose graph is an S–shaped curve (Minh, 2004), and the assumption of speediness; Meaning that the person's failure to answer the item correctly is due to his low ability and not to the speed factor (Park et al., 2019; Almaleki & Alomrany, 2021), these assumptions if were violated, may lead to a poor fit of the model to the data (Hambleton & Swaminathan, 1993; Heene et al., 2016; Temel et al., 2022).

The accuracy of the model in measuring the attribute also depends on the accuracy of estimating its parameters, whether for the item or Individual (Hambleton & Swaminathan, 1985). It is assumed in the IRT that the estimate of the parameters of the item is free from the sample of the subjects, as well as that the estimate of the ability of Individual is free from the items (Harvey, 2016; Al-Tarawnah & Al-Qahtani, 2022). Therefore, psychometricians and educationalists sought to find the best ways to estimate it. One of the methods for estimating the parameters of an item is the Maximum Likelihood Estimation (MLE) method, which depends on the person's response pattern to the item (1 for the correct answer, 0 for the wrong answer), and uses the maximum likelihood function to estimate multiple values of ability, the largest of which is taken to represent the person's estimated ability, where it was found that as the sample size increases, the estimated value of the ability gets closer to its true value (Embertson and Reise, 2000).

Among (MLE) methods are the Joint Maximum Likelihood Estimation (JMLE), in which the ability of Individual and the parameters of the item are estimated together, so that first values of the ability of Individual are assumed, through which preliminary parameters of the items are estimated, and secondly, the parameter of the ability of Individual is estimated based on the parameters of the estimated item so that these two processes are repeated until reaching to stability in the estimation process (Lincare, 1994; Haberman, 2004; Paolino, 2013; Robitzsch, 2021). In the analysis Under the JML procedure item responses are essentially treated as the observational units, and treats both items and abilities as unknown, but fixed model parameter, the model is not identified, which means a unique solution does exist if further constraints are placed on the parameters of the model. For two parameter models like the 2PL, two constraints are necessary: a location constraint, and a scale constraint. The location constraint can be made by constraining either a single propensity or difficulty to some fixed number, or by constraining the average propensity or difficulty to some fixed number (typically zero). The scale constraint can be made by forcing the product of the discrimination parameters to one (Ghosh, 1995; Johnson, 2007).

The Conditional Maximum Likelihood Estimation (CMLE) is used to estimate parameters of the single-parameter logistic model only. Where the association of the probability of an person answering the item accurately is conditional on the number of correct answers of Individual to the test items. The respondent's overall score is a sufficient statistic to calculate approximately an person's ability (Paolino, 2013; Draxler & Alexandrowicz, 2015). In the method Marginal Maximum Likelihood Estimation (MMLE), the marginal maximum likelihood integration is found for the parameters of the items, through the process of integrating the probability density function for the ability parameters, so that the estimation process is carried out in two stages; In the first stage, the Maximization stage, the predicted number of Individual who answer the items correctly and at each level of ability is calculated. In the second stage, the expectation stage, the parameters of the item are used to find the maximum likelihood

function through expectations (Chen & Choi, 2009). The Bayesian method is also used in estimating the parameters of the model when it is difficult to use the MMLE method, especially in the complete answers or not answering all items, as this method assumes the existence of pre-values for the ability of Individual with a normal distribution (Warm, 1978; Chen & Choi, 2009; Almaleki, 2021).

Several approaches for measuring person's ability have arisen, the most important of which are: Maximum Likelihood Estimation (MLE) is the most common and widely used method, in which the parameters are calculated using mathematical processes based on maximizing the parameter to be estimated. It employs the following strategies: Joint Maximum Likelihood Estimation (JMLE), which evaluates both person ability and item parameters. As a result, values for person ability are assumed first, and basic parameters for the items are estimated through them. Second, the person ability parameter is calculated using the estimated item's parameters. These two procedures are repeated until the estimating process reaches a state of stability.

Regarding the Bayesian method, which is favoured to be employed when the test items are completely or incompletely answered. And it depends on its estimation of the parameters on the assumptions of Bayesian estimates, in the sense of using preliminary prior information about the parameters of the item by establishing a prior distribution of ability. Then a random sample is taken from it, and the parameters of the dimensional item are estimated; In the sense of finding a dimensional distribution of the parameters of the item (Posterior Distribution). Finally, in 1985, Choppin developed the paired approach (Pairwise), which is now regarded as one of the modern ways. This method is based on the Thurston model of pair comparisons, which was turned into a practical mechanism to calibrate things by determining the difficulty of items in question banks. Its key benefit is that it does not require lengthy mathematical operations and merely compares two items at a time. It also has the advantage of easily addressing partial data matrices thanks to its numerical methodology, as well as estimating model parameters in the case of missing data. When employed with the pairing approach, the accuracy of calculating the difficulty parameter was better than the ML method with a small sample (Heine & Tarnai, 2015).

Many investigations were undertaken to discover the best ways for evaluating the parameters of the items and Individual. The study (de la Torre & Hong, 2010) sought to determine the effect of sample size on the accuracy of estimating item parameters and ability in tests constructed using (Higher Order Item Response Theory: HO-IRT) in generating responses consisting of (500, 1000). The influence of sample size and test length on the accuracy of estimating item parameters and ability in the test was investigated using the Monte Carlo method on a test of (10, 20) items. The findings demonstrate that the sample size and length of the test influence the estimation of item parameters, with the sample of 1000 examinees and the test 20 items having the lowest standard errors. The accuracy of estimating the ability parameter was not affected by the sample size and was affected by the length of the test, with the standard errors being minimal when using the test of 20 items compared to the length of the test with 10 items.

Chen (2014) also conducted a study with small sample sizes (30, 50, 100, 250) subjects and a 10-item test to evaluate the findings of the Rasch model analysis. The data was analyzed using the Mplus program, and the results showed that when small samples (30, 50) are examined, the standard errors in estimating parameters are higher than when larger samples (100, 250) are studied. Jiang et al (2016) used the MML method to estimate parameters in the flexMIRT software to find the appropriate size to estimate item parameters according to the Multidimensional Graded Response Model, where data were generated with different sample sizes of (500, 1000, 1500, 2000) examinees, and different test lengths of (30, 90, 240) items. When utilizing tests with 30 and 90 items, the findings demonstrate that the smallest sample size that produces valid estimations of item parameters is 500 Individual. When using a test with a length of 240 items, it is necessary to use a sample of at least 1,000 subjects and increasing the sample size to greater than 1,000 subjects does not increase the accuracy of parameter estimation.

To find out the effect of sample size and test length on the accuracy of estimating parameters of item response theory models, (Sahin & Anil, 2017) employed the MMLE in Xcalibre 4.1 software for estimating item parameters. To achieve this, three language tests of different lengths consisting of (10, 20, and 30) items were developed and applied to nine different sample sizes consisting of (150, 250, 350, 500, 750, 1000, 2000, 3000, and 5000) examinees. The results exhibited that the mean of standard errors in estimating the difficulty parameter decreases with the increase in the sample size and that according to the one-parameter model, a sample size of at least 150 subjects can be used with tests consisting of (10, 20, 30) items to accurately estimate the difficulty parameter.

Finch & French (2019) conducted a study to compare the methods for estimating item parameters (JMLE maximum, Bayesian method, pairing method) according to IRT theoretical models with different sample sizes and test lengths. To achieve the objective of the study, different sample sizes were generated consisting of (25, 50, 100, 250, 500, and 1000) subjects and test lengths consisting of (10, 20, 30, 40, and 50) and the R language program was utilized to estimate the parameters of the item. The findings revealed that using the pairing method gives accurate estimates of the item difficulty parameter compared to other methods. It was found that the sample size affects the accuracy of estimating the difficulty parameter of the item, as the sample size decreases, the standard errors decrease in estimating the difficulty parameter.

Almaleki & Alomrany (2021) conducted a study to compare (EAP) estimation method with (MML) method on the accuracy of estimating the items parameters and ability, using the Three Parameter Logistic. an achievement test in chemistry was applied to a sample of (507) students of the third year of secondary school in the "Natural Sciences Course". The study's results revealed that the (MML) method showed a less degree of accuracy in the estimation of the difficulty parameter and the abilities of persons than the (EAP) method. There were no statistically significant differences in the accuracy of the parameter estimation of discrimination and guessing according to estimation method (MML and EAP).

Al-Tarawnah & Al-Qahtani (2022) conducted study using Monte Carlo method of simulation to determine the effect of test length on the estimation of ability parameter in the two-parameter and three-parameter logistic models, using the Bayesian method of expected prior mode and maximum likelihood. The study includes random samples of subjects and of items. Results reveal that with the increase of test length the accuracy of the ability parameter estimation increases in the two- parameter logistic model and three-parameter logistic model according to the maximum likelihood method and the Bayesian method. Results also show that with long and average length tests, the effectiveness is related to the maximum likelihood method and to all conditions of the sample size, whereas in short tests, the Bayesian method of prior mode outperformed in all conditions. The Bayesian method outperforms with respect to the accuracy of estimation at all conditions of the sample size, whereas in long tests the maximum likelihood method outperforms at all different conditions.

By reviewing prior studies, it is concluded that the accuracy of estimating the parameters of the items according to IRT theory models improves as the sample size of the subjects and the length of the test rises, such as the study (de la Torre & Hong, 2010; Zboun, 2013; Finch & French, 2019; Almaleki & Alomrany, 2021; Al-Tarawnah & Al-Qahtani; 2022). According to the difference in sample size and test length, the current study agrees with prior studies in its quest for accuracy in estimating item parameters and the ability to utilize the IRT theory. However, no study has revealed the effectiveness of the JMLE method in estimating item parameters and Individual' ability, thus, this study came to reveal the effectiveness of the JMLE method in estimating the parameters of the item and the ability of Individual according to the 3PL model by generating and analyzing data using the R language software with different sample sizes and length of the test.

Problem Study and Questions

Many researchers in the humanities and social sciences have been interested in using measuring tools such as tests and measures with appropriate

psychometric properties, based on the IR theory because it has advantages that counteract the flaws in traditional measurement theory, the most significant of which is the problem of different item parameters depending on the sample of examinees and the length of the test (Hambleton & Swaminathan, 1985). The use of the item response theory necessitates the best and most optimal method for estimating the parameters of the item and Individual, considering the sample size and length of the test. Therefore, the current study aimed to demonstrate the accuracy of the JMLE method in estimating the parameters of the item and the ability of Individual using the 3PLM model, taking into account the sample size and length of the test. The study sought to do this by answering the following questions:

1) Does the accuracy of estimating item parameters (difficulty, discrimination, and guessing) in the 3PLM estimated using the JMLE method vary with sample size and test length?
2) Does the accuracy of estimating abilities in the 3PLM model estimated using the JMLE method differ with sample size and test length?

Study Significance:

The significance of this study stems from the fact that it examines various methods, both cognitively and theoretically, for estimating the parameters of the items and Individual according to IR theory models, with a focus on the comparison process to determine the effectiveness of the JMLE in estimating the parameters of the item and Individual' ability according to sample size and test length. The current study is important from a practical standpoint since it clarifies for other researchers in the humanities and social sciences the usefulness of the JMLE method in estimating item parameters and person ability based on sample size and test length when they use the item response theory in constructing the various tests.

Study limitations

The current study was limited to using data generated according to the Monte Carlo method, for the responses of Individual with five sample sizes (100, 250, 500, 1000, 2000), on a test of different lengths (20, 40, 60) items according to the 3PL theory model, where the JMLE method was employed to estimate item parameters and person ability.

Methods and Procedures

Data Collecting

Using the Monte Carlo method, the researcher collected responses (100, 250, 500, 1000, 2000) from Individual with abilities that were typically distributed with arithmetic mean 0 and standard deviation 1 $\theta \sim N(0,1)$, on a test of lengths (15, 30, 60, 90) items using the 3PL theory model. The difficulty coefficients for items ranged between -2 and 2 $b \sim U(-2.2)$, and item discrimination coefficients were generated by a~LogNormal distribution (0,0.25). The item guessing coefficients were generated with c~Beta(6,28) distribution where R-Package was employed in the process of data generation and statistical processing.

Statistical analysis

After reviewing previous studies on the accuracy of the estimation of item parameters or the ability of Individual, the researchers applied a variety of criteria to detect the accuracy of parameter estimation, the most prominent of which are: Pearson's correlation coefficient between the expected value and the expected value of the parameter, and the root of mean square errors (RMSE) between the expected value and the parameter's expected value. Two criteria were adopted to assess the accuracy of the item's parameter estimation and person ability.

Results and Discussion

1) **Findings of the first question: "**Does the accuracy of estimating item parameters (difficulty, discrimination, and guessing) in the 3PLM estimated using the JMLE method vary with sample size and test length?**"**

To answer this question, item parameters (difficulty, discrimination, and guessing) were estimated in the 3PLM model estimated using the JMLE method with varying sample sizes and test lengths (see Appendix 1).

The estimated values of the item parameters according to the 3PLM model theory

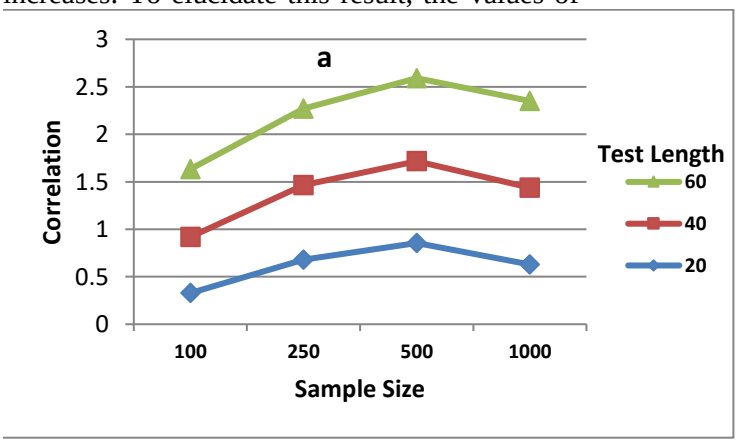
by the JMLE method differ due to the sample size and the length of the test. To detect the difference in the accuracy of estimating the item parameters with the difference in the sample size and the length of the test, the indicator of the correlation coefficient between the expected value and the estimated value was used as shown in Table 2.

Table 2 The values of the correlation coefficients between the expected value and the estimated value of the item parameters estimated by the JMLE method by the sample size and test length

Parameter	Sample size	Test length		
		20	40	60
a	100	0.328	0.591	0.714
	250	0.681	0.784	0.806
	500	0.855	0.861	0.875
	1000	0.628	0.811	0.912
b	100	0.972	0.986	0.988
	250	0.984	0.963	0.978
	500	0.983	0.994	0.948
	1000	0.968	0.988	0.988
c	100	0.021	0.384	0.393
	250	0.284	0.171	0.408
	500	0.492	0.560	0.335
	1000	0.623	0.427	0.686

Table 2 reveals a positive relationship between the expected value of the item parameter and their estimated value, and that when the length of the test increases, the correlation coefficient between the expected value of the item parameter and their estimated value increases. To elucidate this result, the values of the correlation coefficients were represented graphically see (Figure. 1)

Figure 1 The correlation coefficients between the item expected value of the discrimination parameter and its estimated value



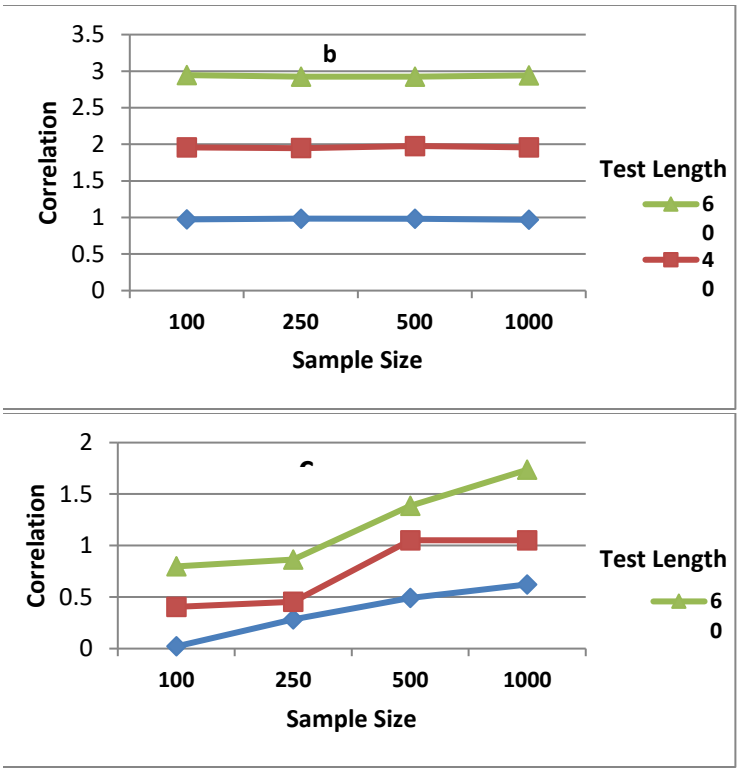


Figure 1 shows that the approximate stability of the value of the correlation coefficient between the expected value and the estimated value of the item difficulty parameter is affected by the increases in the sample size, and the value of the correlation coefficient between the expected value and the estimated value of the discrimination parameter and guessing the item also increases with the rise in the sample size. The RMSE index was also used as shown in Table 3.

Table 3 RSI values of the mean squared errors between the expected value and the estimated value of the item parameters estimated using the JMLE method by sample size and test length

Parameter	Sample size	Test length		
		20	40	60
a	100	0.192	0.156	0.163
	250	0.163	0.100	0.096
	500	0.118	0.126	0.137
	1000	0.152	0.108	0.100
b	100	0.222	0.185	0.132
	250	0.139	0.198	0.142
	500	0.182	0.103	0.255
	1000	0.169	0.118	0.103
c	100	0.061	0.041	0.045
	250	0.056	0.051	0.042
	500	0.041	0.038	0.059
	1000	0.043	0.039	0.038

Table 3 demonstrates that the value of the RMSI decreases. And that when the length of the test increases, the RMSI value decreases, meaning that, the accuracy of guessing the parameters of the items increases as the length of the test increases. To explain this finding, the values of the RMSE rate were represented graphically as indicated in Figure 2.

Fig 2 The values of RMSE

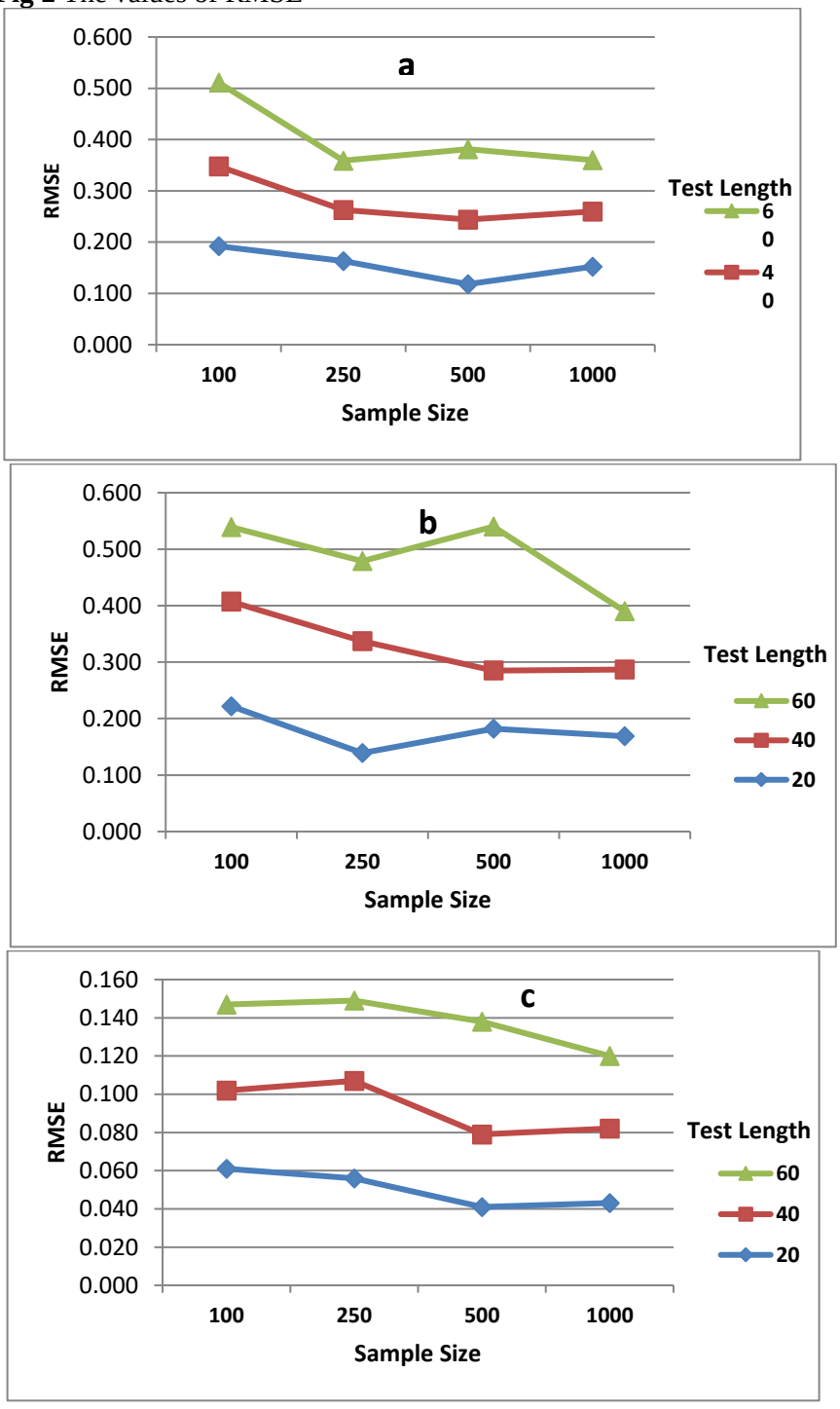


Figure 2 also shows that as the length of the test increases, the value of the RMS error rate increases for the item's parameters, and as the sample size increases, the value of the RMSI decreases for the item's parameters; in other words, increasing the sample size rises the accuracy of estimating the item's parameters.

These findings can be interpreted by the fact that there is a need for the process of estimating the parameters of items in the IR theory with a sufficient number of examinees. As demonstrated previously an increase in the sample size cause an increase in the accuracy of estimating the parameters of the items. This fact was concluded by (Hambleton, 1989) who confirmed that the IR theory needs large sample sizes to obtain accurate estimates of item parameters. These results are consistent with the findings of (Sahin & Anil, 2017), which revealed that the accuracy of the estimation of the item difficulty parameter increases with the increase in the sample size. The results of (Al-Ababneh, 2004) also agree with these results. As they exhibited that the accuracy of the item parameters estimates increases with the increase in the sample size of the examinees. These results are also in agreement with the results of the study by (Huang et al, 2001), which showed that the estimation errors for the difficulty parameter and the discrimination parameter are greater when the sample size of the subjects decreases. That is the mean standard errors of the item parameter estimates are the lowest possible when using a sample size of 1000 examinees.

2) **Findings of the second question: "**Does the accuracy of estimating abilities in the 3PLM model estimated using the JMLE method differ with sample size and test length?**"**

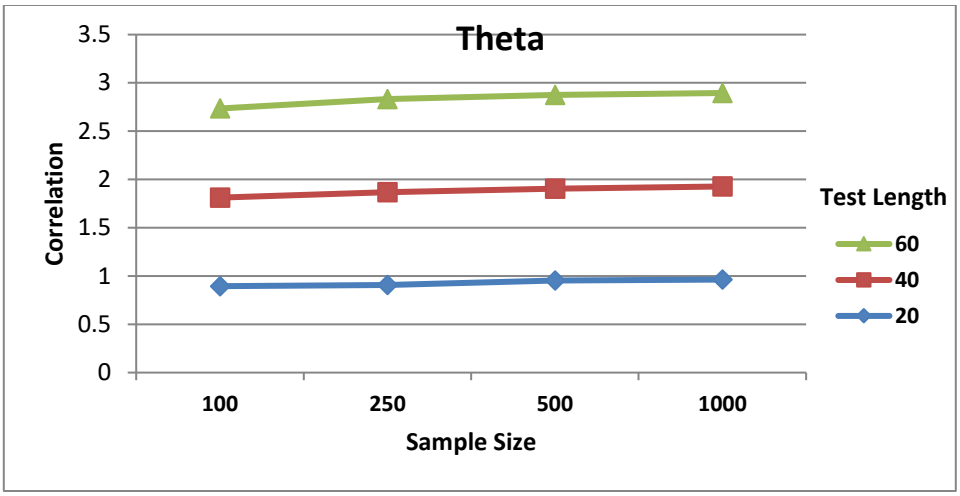
To answer this second question, the indicator of the correlation coefficient between the expected value and the estimated value of the ability of Individual was calculated as indicated in (Table3).

Table 3 The values of the correlation coefficients between the expected value and the estimated value of the person ability parameter estimated by the JMLE method by sample size and test length

Parameter	Sample size	Test length		
		20	40	60
theta	100	0.895	0.917	0.923
	250	0.908	0.959	0.967
	500	0.954	0.951	0.970
	1000	0.963	0.964	0.968

Table 3 reveals a positive relationship between the expected value of the parameters of the item and its estimated value. And that the increase in the length of the test, the correlation coefficient between the expected value of the predicted ability of Individual and its estimated value increases, to justify this, the values of the correlation coefficients were represented graphically in Figure 3 below.

Figure 3 The correlation coefficients between the expected value of the person ability parameter and its estimated value



The index of the value of the root-square index of the mean of the squares of errors of the estimated ability of Individual was also used as shown in Table 4.

Table 4 RMSE rate index value of the person's ability parameter estimated by JMLE method by sample size and test length.

Parameter	Sample size	Test length		
		20	40	60
theta	100	0.464	0.411	0.386
	250	0.423	0.284	0.322
	500	0.316	0.314	0.260
	1000	0.279	0.272	0.264

The value of the RMSI decreases as shown in Table 4. It is clear when the length of the test increases, the RMSI value decreases. This suggests that when the length of the test is increased, the accuracy of estimating an person's ability rises. To justify this finding, the values of the RMSE rate were represented graphically in Figure 4 below.

Fig 4 RMSE rate index values

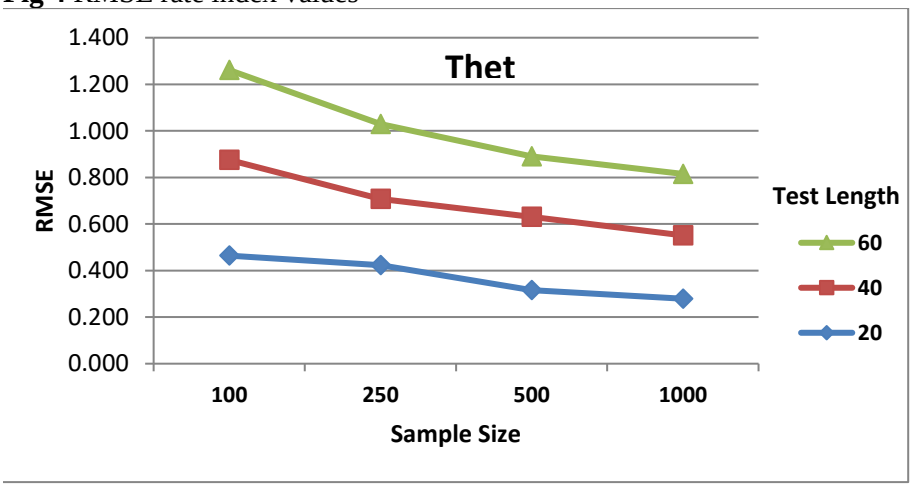


Figure 4 illustrates that as the length of the test increases, the value of the RMSE index of the ability error rate of Individual falls., and that with an increase in the sample size, the value of the

RMSE rate of the ability of Individual decreases, in other words, by increasing the sample size, the accuracy of estimating the ability of Individual raises.

These findings can be explained by the fact that the IR model requires a sufficient number of examinees to estimate the ability of Individual properly because increasing the sample size rises the accuracy of estimating the ability of Individual, and this is confirmed by (Hambleton, 1989) who pointed out that the IR theory needs large sample sizes to obtain accurate estimates of the Individual' ability. These results are in agreement with the results of (Huang et al, 2001), which reported that the errors in estimating the ability of Individual are greater when the sample size is reduced. The mean standard errors of the item parameter estimates are the lowest possible when using a sample size of 1000 examinees. Hambleton & Cook (1983) concluded that the increase in the sample size and the length of the test boosts the accuracy of the estimation of the ability of the examinee.

The findings of the study questions confirmed the significance of the length of the test in increasing the accuracy of estimating the parameters of the items and the ability of the subjects. This finding is consistent with the result (Ababneh, 2004), which showed that the accuracy of the ability parameter estimates increases with the increase in the length of the test.

Recommendations

In light of the aforementioned findings, the researcher recommends the following:

- Employing the JMLE method to estimate item parameters and the person ability for all IR theory models, and with different sample sizes and test lengths.
- Conducting more studies about the estimation of the item parameters and the ability of Individual for all models of IR theory and comparing them with different sample sizes, test lengths and real data.
- Conducting a study to compare the different methods of estimating the item parameters according to the difference in the sample size and length of the test, and with real and generated data.

ACKNOWLEDGEMENTS

The researcher extends her thanks and appreciation to the management of Palestine Technical University Kadoorie for their financial and moral support.

References

1. Almaleki, D. A. (2021). Challenges Experienced Use of Distance-Learning by High School Teachers Responses to Students with Depression. International Journal of Computer Science & Network Security, 21(5), 192-198.
2. Almaleki, D. A., & Alomrany, A. G. (2021). The Effect of Methods of Estimating the Ability on The Accuracy and Items Parameters According to 3PL Model. International Journal of Computer Science & Network Security, 21(7), 93-102.
3. Al-Tarawnah, E. A. W., & Al-Qahtani, M. (2022). The Effect of Test Length on the Accuracy of Estimating Ability Parameter in the Two-and Three-Parameter Logistic Models: Comparison by Using the Bayesian Method of Expected Prior Mode and Maximum Likelihood Estimation. Journal of Educational and Social Research, 12(1), 168-168.
4. Baker, F. B., & Kim, S. H. (2004). Item response theory: Parameter estimation techniques. CRC press.
5. Bulut, O. (2013). Between-Person and Within-Person Subscore Reliability: Comparison of Unidimensional and Multidimensional IRT Models. [PhD thesis]. Minneapolis: University of Minnesota.
6. Chen, J., & Choi, J. (2009). A comparison of maximum likelihood and expected a posteriori estimation for polychoric correlation using Monte Carlo simulation. Journal of Modern Applied Statistical Methods, 8(1), 32.
7. de la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order IRT model approach. Applied Psychological Measurement, 34(4), 267-285.
8. Dohoo, I., & Emanuelson, U. (2021). The use of item response theory models to evaluate scales designed to measure knowledge of,

and attitudes toward, antibiotic use and resistance in Swedish dairy producers. Preventive Veterinary Medicine, 195, 105465.
9. Draxler, C., & Alexandrowicz, R. W. (2015). Sample size determination within the scope of conditional maximum likelihood estimation with special focus on testing the Rasch model. Psychometrika, 80(4), 897-919.
10. Embretson, S. E., & Reise, S. (2000). Item response theory for psychologists. Mahwah: Erlbaum.
11. Finch, H., & French, B. (2019). A comparison of estimation techniques for IRT models with small samples. Applied Measurement in Education, 32(2), 77-96.
12. Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. Statistics & Probability Letters, 23(2), 165-170.
13. Haberman, S. J. (2004). Joint and conditional maximum likelihood estimation for the Rasch model for binary responses. ETS Research Report Series, 2004(1), i-63.
14. Hambleton, R. K., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands.
15. Hambleton,R.k., & Swaminathan,H, & Rogers. H.j.(1993). Fundamentals of Item Response Theory: International Educational and Professional. Publisher Newbury park.
16. Harvey, R. J. (2016). Improving measurement via item response theory: great idea, but hold the Rasch. The Counseling Psychologist, 44(2), 195-204.
17. Heene, M., Kyngdon, A., & Sckopke, P. (2016). Detecting violations of unidimensionality by order-restricted inference methods. Frontiers in Applied Mathematics and Statistics, 2, 3.
18. Johnson, M. S. (2007). Marginal maximum likelihood estimation of item response models in R. Journal of Statistical Software, 20, 1-24.
19. Kim, S. R., Bhang, S. Y., Lim, E. Y., Huh, S., Lee, S. K., Kraus, S. W., & Potenza, M. N. (2021). Reliability, Validity, and Unidimensionality of the Korean Version of the Pornography Craving Questionnaire Based on the Classical Test Theory and Item Response Theory. Psychiatry Investigation, 18(6), 530-538.
20. Linacre, J. M. (1994). Many-Facet Rasch Measurement. Chicago: MESA Press.
21. Liu, T. H., Ho, A. D., Hsu, Y. T., & Hsu, C. C. (2021). Validation of the EQ-5D in Taiwan using item response theory. BMC public health, 21(1), 1-9.
22. Mutiawani, V., Al Misky Athaya, K. S., & Subianto, M. (2022). Implementing Item Response Theory (IRT) Method in Quiz Assessment System. TEM Journal, 11(1), 210-218.
23. Ogunsakin, I. B., & Shogbesan, Y. O. (2018). Item response theory (IRT): a modern statistical theory for solving measurement problem in 21st century. International Journal of Scientific Research in Education, 11, 627-635.
24. Osterlind, S. J., & Wang, Z. (2017). Item response theory in measurement, assessment, and evaluation for higher education. In Handbook on measurement, assessment, and evaluation in higher education (pp. 191-200). Routledge.
25. Paolino, J. P. N. (2013). Penalized joint maximum likelihood estimation applied to two parameter logistic item response models (Doctoral dissertation, Columbia University).
26. Park, J. Y., Cornillie, F., Van der Maas, H. L., & Van Den Noortgate, W. (2019). A multidimensional IRT approach for dynamically monitoring ability growth in computerized practice environments. Frontiers in psychology, 10, 620.
27. Philip, A., & Ojo, B. O. (2017). Application of item characteristic curve (ICC) in the selection of test items. British Journal of Education, 5(2), 21–41.
28. Robitzsch, A. (2021). A comprehensive simulation study of estimation methods for the Rasch model. Stats, 4(4), 814-836.
29. Sijtsma, K., & Junker, B. W. (2006). Item response theory: Past performance, present developments, and future expectations. Behaviormetrika, 33(1), 75-102.
30. Świst, K. (2015). Item analysis and evaluation using a four-parameter logistic model. Edukacja, 134(3), 79-99.

31. Temel G, Machunsky M, Rietz C and Okropiridze D .(2022). Investigating Subscores of VERA 3 German Test Based on Item Response Theory/Multidimensional Item response Theory Models. Front. Educ. 7:801372. doi: 10.3389/feduc.2022.801372.

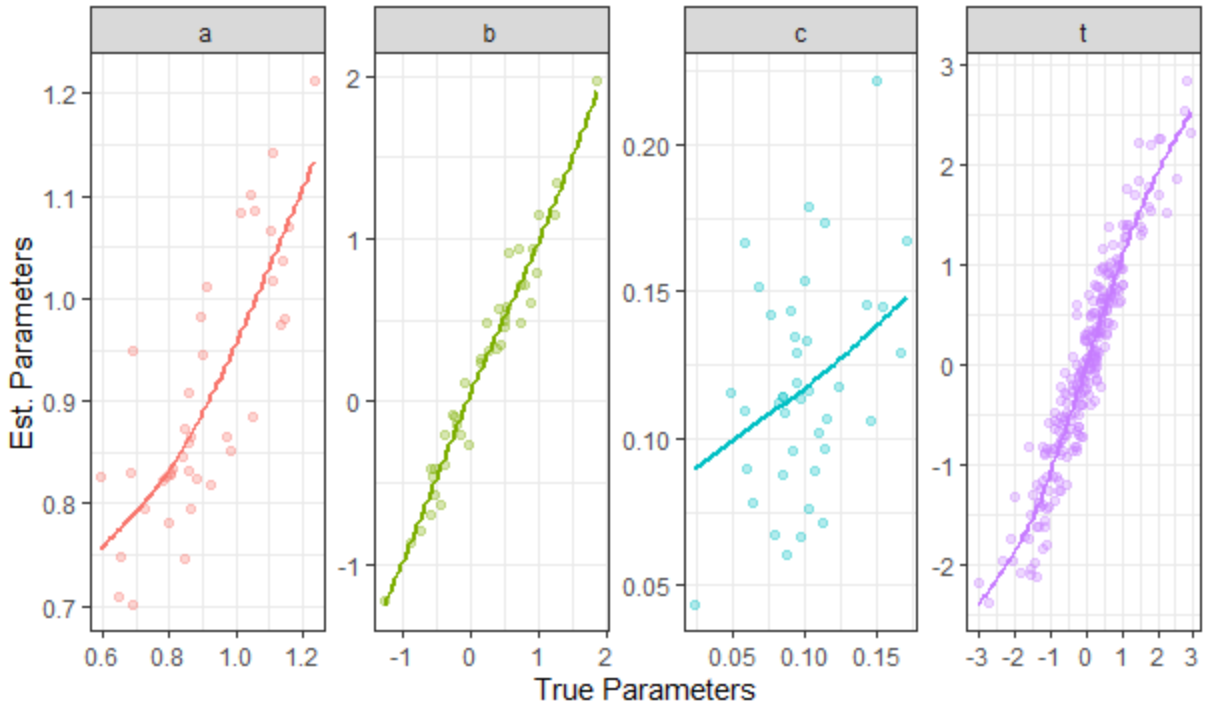
Appendix I Item parameter values (difficulty, discrimination, and guessing) in the 3PLM model estimated by the JMLE method with varying sample size and test length

Item No.	100			250			500			1000		
	a	b	c	a	b	c	a	b	c	a	b	c
1	1.2144	-0.9030	0.0496	1.0317	0.5929	0.1602	1.2619	-0.8162	0.1181	0.8762	-0.2241	0.0930
2	0.7685	-1.0672	0.0932	0.8916	-0.5505	0.0621	0.7291	-1.0587	0.0822	1.1396	-1.1579	0.1064
3	0.7511	0.2821	0.1346	1.1366	0.6460	0.0976	1.1089	-0.5625	0.0582	1.0091	-0.3434	0.1049
4	1.0271	-0.6262	0.0498	0.9535	-0.0678	0.0849	1.0962	-0.9556	0.1484	1.1867	0.5472	0.0541
5	0.9991	-0.7807	0.1860	0.7815	0.5350	0.1008	1.2214	-0.3490	0.1218	0.9655	-0.1040	0.0886
6	0.9136	0.9832	0.0838	0.9330	0.3495	0.0974	0.9410	-0.1572	0.0584	0.9614	-0.9548	0.1150
7	0.7476	0.1286	0.0404	1.0302	1.8022	0.1194	0.7100	0.8340	0.0816	0.7906	0.3778	0.1224
8	0.9605	-0.8600	0.1479	0.6159	-0.0016	0.1169	0.6946	0.5814	0.0767	0.7107	1.3675	0.0438
9	0.9933	-0.3876	0.1034	0.9621	0.6713	0.1111	0.8564	-0.3329	0.1407	0.8079	-0.0195	0.2305
10	1.2399	-0.3407	0.1019	0.7878	0.0259	0.0325	0.9601	0.1362	0.0603	0.6363	-1.4420	0.1722
11	1.4185	0.2813	0.1454	1.1710	-0.7838	0.0627	1.1146	0.6379	0.0638	0.9535	0.6670	0.0703
12	0.7020	1.5153	0.0474	0.6637	1.0602	0.0228	0.8668	-0.1379	0.0996	0.6513	-0.0969	0.1137
13	0.7964	-0.6514	0.1011	0.8434	-0.3839	0.0961	1.1367	0.7490	0.2309	0.8046	-0.5648	0.0561
14	0.8962	0.3583	0.1195	1.1502	1.4651	0.1179	0.9565	0.4792	0.0761	0.6922	-0.1798	0.0949
15	0.9099	0.0574	0.0585	0.9940	0.0067	0.0817	0.8012	0.4184	0.1328	0.7364	0.6658	0.0624
16	0.7221	0.0913	0.1125	0.8042	1.1368	0.1316	0.9754	1.1242	0.0959	0.9097	-0.6927	0.1128
17	0.7559	0.1003	0.1264	1.3249	-0.4546	0.1066	0.7813	0.8566	0.1823	0.9067	-0.4499	0.0538
18	0.9129	-1.5885	0.0961	0.8191	-0.8557	0.1222	1.0301	-0.1668	0.0846	0.9170	-0.4149	0.0634
19	1.0525	0.2861	0.0699	0.8907	-0.6129	0.0294	1.1702	-0.2133	0.0649	0.8884	0.2687	0.0851
20	0.8081	0.1054	0.0454	0.9614	-0.9342	0.0386	0.8409	1.0557	0.0945	1.1207	-1.4220	0.0881
1	1.0967	0.5220	0.0865	1.1061	-0.5111	0.0988	0.7436	1.1851	0.0660	0.9145	0.9963	0.0790
2	0.9255	-0.5729	0.0668	0.8580	0.8926	0.1130	0.7383	-0.7190	0.0402	0.8762	-0.6249	0.0550
3	0.7210	0.4114	0.0764	0.8031	0.7514	0.1059	0.6913	-0.6409	0.0428	0.7418	-0.1489	0.0977
4	1.1852	0.4859	0.1677	1.0562	0.5139	0.0786	1.2205	-0.4951	0.1098	0.9570	0.0671	0.1330
5	1.0804	-0.8203	0.0625	0.6522	-0.8602	0.0673	0.8027	-0.7261	0.0668	0.9457	-0.7788	0.0852
6	0.6634	-1.0452	0.0790	1.1584	1.2322	0.1018	1.3619	-0.1226	0.0795	0.7469	-1.5160	0.1036
7	0.9617	-1.3116	0.0626	1.0507	-0.5641	0.1663	0.7412	0.1366	0.0830	0.8370	-0.0685	0.0816
8	1.0625	-0.4867	0.0831	0.8567	0.2775	0.0972	0.8656	0.4980	0.0929	0.8496	-0.4277	0.1075
9	1.0135	0.4517	0.0530	1.1070	0.5019	0.1126	0.4747	-0.4346	0.1052	0.8648	-0.8424	0.0703
10	1.0196	0.1434	0.0355	0.8126	-0.3816	0.1007	1.0110	-0.2612	0.0969	0.8505	0.4959	0.0491
11	0.8846	-0.1882	0.1386	1.1122	0.7171	0.0228	1.0246	-0.6008	0.0885	0.7443	0.7227	0.0905
12	0.7445	-0.4054	0.0988	0.9019	0.5509	0.1537	1.0142	0.0276	0.1305	0.7248	0.8580	0.1494
13	1.1519	0.7490	0.0539	1.1473	0.1458	0.1705	0.7585	-0.5719	0.0903	0.7527	-0.8963	0.0775
14	1.2284	0.2564	0.0789	1.1322	-0.7374	0.0479	0.7658	-0.6570	0.1850	0.6617	-0.9884	0.1542
15	0.9602	-1.4794	0.0913	0.8012	0.9704	0.0846	1.1387	0.5176	0.0586	0.7151	0.2452	0.1036
16	0.6633	-0.2287	0.1265	1.0426	-0.1477	0.1029	1.0260	-0.2402	0.0523	0.7226	-0.5921	0.1469
17	1.0288	-0.0646	0.0859	0.7228	-0.5452	0.0753	0.7870	0.0470	0.0808	1.1445	-0.1399	0.0478

Item No.	100			250			500			1000		
	a	b	c	a	b	c	a	b	c	a	b	c
18	0.8339	1.4014	0.0889	0.8445	0.8121	0.0872	1.1197	0.8414	0.1912	0.8054	-0.7731	0.0781
19	1.1090	-0.5257	0.1032	0.6884	0.5786	0.1021	0.8388	0.3544	0.0845	0.9793	1.0649	0.1103
20	0.8734	0.5524	0.1152	0.8952	0.9220	0.1092	0.9118	0.4733	0.1292	1.0940	0.5644	0.1229
21	1.2417	-1.2027	0.1318	0.7812	0.4282	0.1498	0.7926	0.1642	0.1080	0.8235	0.3895	0.0393
22	0.9375	0.4972	0.1612	0.8395	0.5212	0.0631	1.0824	1.7959	0.1315	0.8660	0.0656	0.1543
23	0.8625	1.0246	0.1823	0.5957	-0.2520	0.0890	1.0726	0.5071	0.1528	0.9422	0.4443	0.0722
24	0.8094	-0.3139	0.1071	0.9243	-0.3799	0.0576	1.1446	-0.8191	0.1293	0.8351	0.1863	0.1276
25	0.8920	-0.5879	0.1758	0.9698	0.3948	0.0844	0.8567	-1.3347	0.1104	0.8820	-1.7480	0.0646
26	1.0531	0.9743	0.0696	0.7949	0.4603	0.1423	0.7701	-0.0471	0.1091	0.9354	-0.1016	0.1631
27	1.0047	-0.8330	0.0743	0.9127	-0.2010	0.0592	1.0242	-1.0988	0.1774	0.7998	-0.0328	0.0652
28	0.6046	-0.8838	0.0556	1.2333	-0.4170	0.0911	0.8828	-0.0928	0.2322	0.8609	-0.4192	0.1416
29	1.1414	0.4395	0.0529	1.1389	1.8754	0.0971	1.1333	-0.2556	0.0627	1.1626	-0.1652	0.0956
30	0.9078	-0.6508	0.1018	0.6874	0.2465	0.0580	1.1918	-1.0883	0.0811	0.6699	-0.5306	0.1200
31	1.1322	-0.1438	0.0962	0.6858	-0.5827	0.1238	0.8755	1.1267	0.1184	0.8295	-0.3373	0.0883
32	0.8056	-0.3152	0.0823	0.8639	1.2659	0.0815	0.9045	0.2675	0.1122	1.1519	1.4865	0.1308
33	0.8936	0.9442	0.1467	1.0165	-1.2520	0.0858	0.8923	0.1123	0.1183	0.7737	0.9900	0.0638
34	0.9049	0.2300	0.0935	0.8592	0.7532	0.1142	0.9106	0.7077	0.0883	1.4500	-0.3724	0.0421
35	0.8394	-0.2349	0.0905	0.7911	1.0254	0.0931	0.8725	-1.0784	0.0913	0.6941	-1.2861	0.0655
36	0.9801	0.3898	0.1104	0.8661	-0.5207	0.0932	0.7633	0.3084	0.0721	1.0637	-0.4867	0.1167
37	0.7300	0.4770	0.0844	0.6490	-0.0862	0.1127	1.2499	-0.7974	0.0924	0.8530	1.8029	0.0726
38	0.7885	-0.2866	0.1563	0.8820	0.1598	0.0838	0.7766	-0.3974	0.0776	0.8492	0.5849	0.1065
39	0.9715	-0.0290	0.1088	0.9859	-0.2283	0.0942	0.9799	-1.2682	0.0715	1.1610	0.6064	0.0608
40	0.8926	-1.0299	0.0520	0.8467	-0.0234	0.1455	0.9438	0.9610	0.1831	1.1246	-1.2278	0.0484
1	1.0101	-0.4370	0.0915	0.7686	0.1358	0.0551	1.0943	-0.6178	0.1561	0.6591	0.9579	0.0278
2	0.9129	-1.0258	0.0935	0.6917	-0.1639	0.0758	1.0602	0.0970	0.1018	0.8434	-0.2137	0.1159
3	0.9366	-1.0890	0.1091	0.9091	1.1368	0.0693	0.9545	0.7541	0.0604	0.6126	-0.3758	0.1873
4	0.9098	-0.0403	0.0965	0.9091	0.4000	0.0361	0.9903	0.8962	0.1394	0.6766	-0.0340	0.1106
5	0.7935	-0.2837	0.1031	0.7449	0.2113	0.0737	0.6744	-0.1561	0.1426	1.0361	-0.7183	0.0311
6	1.0137	-0.0396	0.0824	0.7805	0.8910	0.0644	0.8757	1.3117	0.0999	1.0017	1.1892	0.1217
7	0.9295	1.1718	0.1151	0.7042	1.8855	0.1583	0.6372	0.5983	0.1237	1.0445	0.1768	0.0711
8	1.2125	0.4030	0.0768	0.8346	-0.3807	0.1090	1.0755	-0.4942	0.0343	1.2207	-0.0408	0.0897
9	0.7631	0.4759	0.0652	0.7666	0.0527	0.0347	0.9330	0.1014	0.0942	1.1630	-0.0558	0.1342
10	0.6975	-0.0646	0.1446	0.7950	0.5319	0.1442	1.0421	-1.4763	0.1061	1.3509	1.0325	0.1167
11	0.9050	0.4400	0.0798	0.9246	-1.1353	0.1188	0.7378	-0.0130	0.0889	0.9052	-0.1638	0.1288
12	0.8058	0.7526	0.1195	0.8846	0.2834	0.1111	1.1899	1.5820	0.0911	0.7224	-0.2437	0.0429
13	1.4125	-0.4212	0.0768	0.8351	-0.1907	0.0454	0.6776	0.4097	0.1100	1.1445	0.5091	0.1425
14	0.8755	0.6415	0.0890	1.2006	0.4299	0.0983	0.6719	-0.8028	0.1536	0.8769	0.8984	0.0945
15	0.9698	-0.8494	0.0295	0.9943	-0.2089	0.2212	0.8293	0.2204	0.0949	0.6550	0.6420	0.1042
16	0.7950	0.4956	0.1565	0.9678	1.3025	0.0926	1.1920	0.3492	0.1416	0.8745	1.1894	0.0628
17	0.9478	0.9482	0.0588	0.9127	-0.2745	0.1387	1.0431	1.6725	0.0751	0.6222	0.0905	0.0540
18	0.7889	-0.5809	0.0225	0.9807	0.9996	0.1810	0.8103	0.8061	0.0267	1.0403	0.9236	0.0256
19	1.2769	-0.4943	0.0954	1.2837	-0.3504	0.1429	0.6505	-0.1947	0.0912	0.8142	0.7821	0.1417
20	1.1824	0.3939	0.1192	1.0904	-0.0893	0.0812	0.7802	0.0757	0.0591	0.8076	0.1504	0.1025

Item No.	100			250			500			1000		
	a	b	c	a	b	c	a	b	c	a	b	c
21	0.9225	-0.1067	0.1227	0.7174	0.8402	0.2125	1.3471	1.4489	0.0435	0.6594	0.5671	0.1899
22	0.7381	-0.5710	0.0212	0.7889	0.1509	0.0420	0.9852	-0.0392	0.0763	0.8849	-0.8071	0.1231
23	0.9098	0.5563	0.0378	1.1734	-0.1824	0.0429	0.6776	-0.3276	0.1973	0.8802	0.6836	0.1769
24	0.9000	0.1033	0.0542	1.4382	-0.0145	0.0773	1.2594	0.3527	0.1230	1.0430	1.5578	0.0423
25	0.5451	-0.0272	0.0930	1.2124	0.3344	0.0530	0.9698	1.2136	0.1064	0.7092	-0.1534	0.1109
26	1.1235	-0.4337	0.1368	0.6905	-1.0798	0.1079	1.0140	-0.6001	0.1115	0.8613	0.1936	0.1497
27	0.9056	-0.8878	0.0476	1.0442	0.7405	0.0511	1.1307	-0.1969	0.0906	1.0094	-0.0317	0.0661
28	0.7771	-0.0226	0.1592	1.1711	0.6479	0.1271	0.7810	0.8326	0.1408	0.8241	-1.0847	0.2105
29	0.7108	-0.1460	0.0987	0.6859	-0.2057	0.0619	1.0695	0.2321	0.1084	0.7575	-0.3005	0.1753
30	1.0943	-0.4686	0.0518	0.8954	1.7738	0.1883	1.0655	-0.3800	0.0676	0.5697	-0.0385	0.0498
31	0.8353	0.1151	0.0422	0.9425	-0.1526	0.0724	0.8052	-0.2662	0.1066	0.8865	-0.1299	0.0521
32	0.8190	1.2893	0.1561	0.9689	-0.0151	0.1108	0.9309	-0.5496	0.1781	0.6824	-0.0594	0.1175
33	0.7143	-0.2754	0.0904	0.6952	0.8645	0.2184	0.8563	-0.1282	0.1513	0.6102	0.3862	0.1392
34	0.9132	-0.6980	0.1267	1.1326	-0.2026	0.1177	0.7726	-0.0390	0.1195	0.9556	-0.0865	0.0515
35	1.0432	0.8309	0.1609	0.8533	-0.4943	0.0791	0.9517	0.1450	0.1327	0.6095	0.5544	0.0248
36	0.9287	0.2195	0.0505	0.8037	0.4525	0.1463	0.9931	-0.0105	0.0562	0.8634	-0.2407	0.0881
37	1.1979	-0.3281	0.1138	1.1365	-0.4316	0.1108	0.6357	0.7444	0.0290	0.9374	-1.2318	0.0933
38	0.5976	0.5346	0.1131	0.9883	-0.0930	0.1657	0.8534	0.3459	0.1260	0.7818	0.1929	0.0670
39	0.8798	-0.8779	0.1080	0.8466	1.1689	0.0842	0.9634	-0.1106	0.2014	1.0288	0.2641	0.1008
40	0.8033	-0.3013	0.0913	0.8888	-0.5604	0.1335	0.8791	0.8611	0.1456	0.8422	0.5943	0.1467
41	0.9408	-0.3674	0.0339	0.8958	-0.8682	0.0350	0.8446	0.9566	0.0510	0.5739	0.1142	0.0533
42	0.8952	-0.7988	0.0174	1.1061	0.5107	0.1420	1.0127	-0.5125	0.1181	0.6682	-0.6522	0.0851
43	0.9091	0.1044	0.0464	1.2090	-1.1503	0.1483	1.1129	1.4913	0.0940	0.8911	-0.8077	0.0634
44	1.0976	-0.4692	0.0733	0.7622	0.2780	0.1074	1.3419	-0.0596	0.0973	0.8413	-0.8712	0.2409
45	1.1539	0.1021	0.0498	0.8865	-0.8696	0.0316	0.8366	-0.9338	0.1272	0.9207	-0.3475	0.0799
46	0.6252	-0.0213	0.0512	0.8644	-0.1315	0.0652	1.0781	-0.9295	0.0352	0.7242	-0.2083	0.1321
47	0.6657	0.0576	0.1193	0.7997	-1.0348	0.1137	0.8821	0.5100	0.0792	0.5863	-0.1266	0.1002
48	0.7439	-0.1017	0.0894	0.8063	0.4486	0.0648	0.9982	-0.7026	0.0519	0.6406	0.2464	0.1100
49	0.9524	0.6342	0.0614	0.6341	-0.4627	0.0827	0.7880	0.6341	0.1037	1.2903	0.3654	0.1053
50	0.7938	0.7848	0.1419	1.4424	0.2890	0.1565	0.7959	-0.6788	0.1259	0.8580	0.7142	0.0739
51	0.6962	-0.0234	0.1956	1.0958	-0.4842	0.0621	1.0072	-0.0810	0.1333	0.9531	-0.7901	0.0618
52	0.8049	-0.0248	0.0729	1.3761	0.4879	0.0664	0.7047	-1.1168	0.0770	0.6610	-0.0660	0.1802
53	0.8731	0.3249	0.0506	0.8281	0.7158	0.0466	0.7815	0.2400	0.0849	0.7889	0.2107	0.0598
54	1.0308	-0.1282	0.0650	1.0286	-0.8464	0.0540	0.9526	-0.2936	0.0916	1.3026	-1.2610	0.0329
55	0.7439	-0.6590	0.0560	0.8813	0.4905	0.0604	0.8435	0.1716	0.0835	0.9267	0.2385	0.1393
56	0.7889	0.2380	0.0849	0.8151	-0.0762	0.1047	0.8823	-0.7305	0.0672	0.8621	0.0130	0.1001
57	0.9705	0.3297	0.1204	1.1120	0.4595	0.0550	0.8525	-0.7352	0.1165	0.8061	-0.0280	0.2080
58	0.7113	0.4861	0.1799	0.8995	0.4348	0.0386	0.8706	-0.4173	0.1732	0.9803	1.2029	0.0689
59	0.6524	0.0092	0.0605	0.9140	-0.5780	0.1244	0.8322	0.9210	0.0623	1.3428	0.3578	0.1284
60	0.9504	-0.3589	0.1345	0.7432	0.0584	0.1098	0.9812	0.8349	0.1027	0.7698	0.4218	0.0920

Appendix 2 The correlation coefficients between the expected value of the person ability parameter (t), Discriminant(a), difficulty(b), guessing(c), and its estimated value(with sample size=10, test length=20)


a
b
c
t
Est. Parameters
True Parameters
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.8
1.0
1.2
2
1
0
-1
-1
0
1
2
0.20
0.15
0.10
0.05
0.05
0.10
0.15
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3