International Journal of Assessment Tools in Education
2021, Vol. 8, No. 2, 423–453
https://doi.org/10.21449/ijate.790289
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

A Guide for More Accurate and Precise Estimations in Simulative Unidimensional IRT Models

Fulya Baris Pekmezci [1,*], **Asiye Sengul Avsar** [2]

[1]Bozok University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education, Yozgat, Turkey
[2]Recep Tayyip Erdoğan University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education, Rize, Turkey

ARTICLE HISTORY
Received: Sep. 04, 2020
Revised: Mar. 30, 2020
Accepted: Apr. 11, 2021

Keywords:
Monte carlo simulation study,
Replication,
Unidimensional item response theory models,
Bias estimation,
Type I error inflation.

Abstract: There is a great deal of research about item response theory (IRT) conducted by simulations. Item and ability parameters are estimated with varying numbers of replications under different test conditions. However, it is not clear what the appropriate number of replications should be. The aim of the current study is to develop guidelines for the adequate number of replications in conducting Monte Carlo simulation studies involving unidimensional IRT models. For this aim, 192 simulation conditions which included four sample sizes, two test lengths, eight replication numbers, and unidimensional IRT models were generated. Accuracy and precision of item and ability parameter estimations and model fit values were evaluated by considering the number of replications. In this context, for the item and ability parameters; mean error, root mean square error, standard error of estimates, and for model fit; M_2, $RMSEA_2$, and Type I error rates were considered. The number of replications did not seem to influence the model fit, it was decisive in Type I error inflation and error prediction accuracy for all IRT models. It was concluded that to get more accurate results, the number of replications should be at least 625 in terms of accuracy of the Type I error rate estimation for all IRT models. Also, 156 replications and above can be recommended. Item parameter biases were examined, and the largest bias values were obtained from the 3PL model. It can be concluded that the increase in the number of parameters estimated by the model resulted in more biased estimates.

1. INTRODUCTION

To make sense of human behavior, individuals need to be observed and evaluated accurately. According to these evaluations, it is important to make decisions about individuals or to direct them towards their needs in a true way. Therefore, the psychometric properties of the measurement tools used for evaluations must be at satisfactory levels.

Test theories are used to assess the psychometric properties of measurement tools. Test theories can be considered as a study area where research is conducted to investigate problems affecting

*CONTACT: Fulya BARIŞ PEKMEZCİ fulyabaris@gmail.com Bozok University, Department of Educational Sciences, Measurement and Evaluation in Education, Yozgat, Turkey

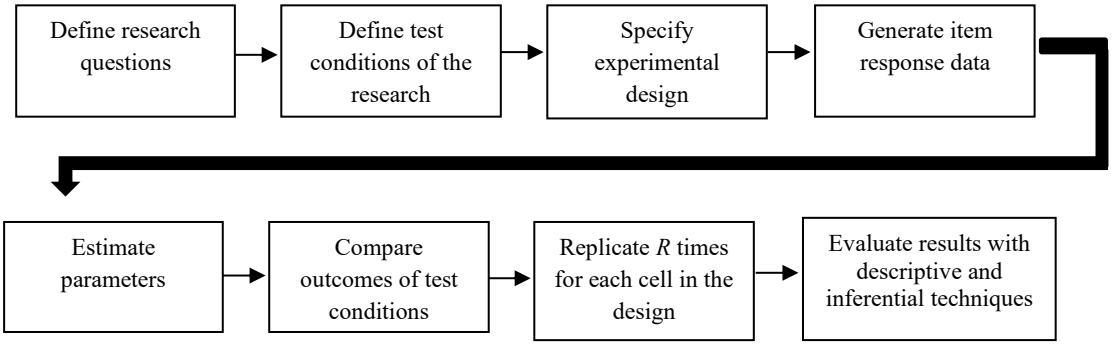
ISSN-e: 2148-7456 /

psychological measurements and to achieve valid and reliable measurement results by trying to reduce these problems as much as possible (Crocker & Algina, 1986). In the literature, the classical test theory and item response theory (IRT) are the most studied theories in the psychometric area.

Human behavior is the main subject of social sciences. It is very important to measure human characteristics, which are very variable, validly, and reliably. The measurement of human behavior is different from the measurements made in natural science. Ideal laboratory conditions are created to achieve the most accurate results in natural science, but it is very difficult to apply these in social sciences. One of the best ways to achieve accuracy in social sciences is through simulation studies. Simulation studies have been used since 1900 as a solution to statistical problems (Harwell et al., 1996).

IRT has strong assumptions that differ according to dimensionality, linearity, or scoring type (McDonald, 1982). In cases where the IRT assumptions are not met, the results of the analysis and estimates will be inaccurate. Monte Carlo (MC) simulation studies provide solutions to the problems that can be encountered by creating ideal data sets that meet the assumptions required for IRT (Han, 2007). MC simulation studies are used for many purposes such as the evaluation of new parameter estimation procedures, comparison of different item analysis programs, and parameter estimation in multidimensional data sets (Harwell, 1997). MC studies perform statistical sampling experiments via computers for solutions to statistical problems (Mundform et al., 2011). How MC studies are structured in IRT (Harwell et al., 1996) is shown in Figure 1.

Figure 1. *Steps of a MC Simulation Study in IRT.*



The MC process starts with defining the research question, as seen in Figure 1. When the research questions defined in psychometry are related to theoretical studies, especially include comparing different conditions at the same time, a simulation study is inevitable in obtaining the appropriate data sets. Then, it is important to define the test conditions of the research. These conditions consist of dependent variables such as item and ability parameters, and independent variables like test lengths, sample size, or distribution of the sample. After specifying the experimental design, the item response data are generated by the IRT model which is chosen by the researchers. Item and ability parameters are estimated from the generated data sets. Results obtained from different test conditions are compared. This process is replicated *R* times and all outcomes are evaluated for answering the research questions using descriptive and inferential techniques.

One of the important issues to be considered in MC simulation studies is the number of replications. With insufficient replications, estimations can be inaccurate (Mundform et al., 2011). Besides, replication is often confused with iteration in the literature. Hence, it is important to clarify the difference between replication and iteration in simulation studies. Iteration is defined as a statistical routine. This routine starts with the first estimate and

continues until a satisfactory estimate, which means the convergence criterion is met, is obtained by working with some statistical rules on this first estimation (Fu, 2019; Thompson, 2004; 2006).

Iterations are needed for convergence of statistical algorithms (Hair et al., 2019). Some iteration algorithms which are used for parameter estimation in IRT are: the Broyden-Fletcher-Goldfarb-Shanno Algorithm, the Bisection Method, the Expectation-Maximization Algorithm, Fisher Scoring, the Gibbs Sampling Algorithm, the Markov Chain Monte Carlo Algorithm, the Newton-Gauss Algorithm, and the Newton-Raphson Algorithm (Cai & Thissen, 2014; Chalmers, 2012; Hanson, 1998; Patsias et al., 2009; Tavares et al., 2004; Thompson, 2009; van der Linden, 2018; Weismann, 2013).

As for replication, this is defined as the repeated administration of an experiment with selected changes in parameters or test conditions being made by the researcher (Hair et al., 2019; Rubinstein, 1981). Replications give an estimate of the stability of the predictions made in simulation studies (Feinberg & Rubright, 2016). Because the number of replications affects the accuracy and reliability of parameter estimates (Feinberg & Rubright, 2016), it is stated that the number of replications is an important factor for statistical results (Kleijnen, 1987; Rubinstein, 1981). These estimations are directly related to the implications to be reached in simulation studies. When conducting a MC simulation study, it is important to answer the question of how many replications are needed for accurate estimations. So, the number of replications should be determined carefully by the researchers. Within the context of unidimensional IRT models, various studies that are conducted on the MC method with a different number of replications are given in Table 1.

Table 1. *Literature review about the number of replications for unidimensional IRT models.*

Studies	Number of Replication
Sheng & Wikle, 2007	10
Roberts et al., 2002	30
Sen et al., 2016	50
Crişan et al., 2017; Lee et al., 2017; Park et al., 2016; Yang, 2007; Zhang, 2008	100
Matlock Cole & Paek, 2017	200
Feinberg & Rubright, 2016	250
Matlock & Turner, 2016	500
Ames et al., 2020; Reise et al., 2011	1000
Baldwin, 2011; Mundform et al., 2011	5000
Babcock, 2011	10000

As is seen from Table 1, the different number of replications ranges between 10 and 10000. It is usual for a different number of replications to be made in varying test conditions for accurate parameter estimations by different IRT models. However, it is not clear what the appropriate number of replications should be under varying test conditions for unidimensional IRT models. In addition, it is important to determine a sufficient number of replications according to test conditions that are specified by the researchers. Although simulative studies provide convenience to theoretical studies, they are time-consuming processes.

To establish a rule for what ideal replication number should be, Feinberg and Rubright (2016) had provided a formula about replication number, which is given in Equation 1:

$$\sigma_M = \frac{\hat{\sigma}}{\sqrt{R-1}} \quad \text{(Equation 1)}$$

where $\hat{\sigma}$ is the standard deviation of the estimated parameter across replications and R is the number of replications and σ_M is the SE of the mean.

According to their formula, they suggested calculating the ideal number of replications by using the standard deviation of the estimated parameters across replications. To determine the ideal replication number, firstly, the researchers must replicate data, and secondly, the ideal replication number must be calculated according to replicated samples' standard deviation. Starting replication number will be the determiner of the ideal replication number. This seems a time-consuming process. Because, firstly, data need to be replicated, and then the ideal replication number must be calculated. Doing more replication will result in a smaller standard deviation of replicated samples or vice versa. Hence, the calculation of ideal replication number according to Feinberg and Rubright (2016) will tend to be smaller due to using that standard deviation. Large standard deviations will recommend more replications. Lastly, there is no exact rule about what the ideal standard deviation of replicated samples should be (see for details Feinberg & Rubright, 2016). Therefore, using Equation 1 does not seem very practical.

In this study, the number of replications required for the most accurate parameter estimations in various sample sizes and test lengths according to unidimensional IRT models (1PL model, 2PL model, and 3PL model) was determined.

The purpose of the current study is to develop guidelines for the adequate number of replications in conducting MC simulation studies involving unidimensional IRT models with different test conditions. Based on this purpose, answers to the following research questions were sought:

1. How are the estimations of item parameters obtained from varying sample sizes and test lengths affected by varying numbers of replications?
2. How are the estimations of ability parameters obtained from varying sample sizes and test lengths affected by varying numbers of replications?
3. How are the estimations of model fit obtained from varying sample sizes and test lengths affected by varying numbers of replications?

2. METHOD

2.1. Study Design Factors

The purpose of this study is to develop guidelines for the adequate number of replications in conducting MC simulation studies involving unidimensional IRT models with different test conditions. According to this aim, different sample sizes and test lengths were studied to determine the adequate number of replications to obtain more accurate and precise estimations.

In line with this purpose, firstly, studies which implemented unidimensional IRT models and MC simulation studies were reviewed. According to the literature review (Baldwin, 2011; Mundform et al., 2011), 5000 was selected as a starting replication number for this study. In determination of other numbers of replications, the method which Preecha (2004) implemented in his study was used. Considering this method, if the bias difference between two consecutive replication numbers is large, this interval should be halved, and the analysis should be repeated. If not, then the last replication number should be halved, the analysis should be repeated, and the bias statistics should be calculated.

After determining the maximum replication number as 5000, bias analyses were performed. Half of the 5000 replications were taken, and the analyses were re-run for 2500 replications. This process was performed until the number of replications was 78. Additionally, the minimum number of replications was determined as 20. In some nonparametric IRT studies, 20 is used as the minimum number of replications (Şengül Avşar & Tavşancıl, 2017; van Onna, 2004). Therefore, in this study, the adequacy of 20 replications was also tested.

Within the scope of this study, a literature review was also done for the test lengths and sample sizes which are given in Table 2. In IRT studies, there are no exact rules for adequate sample sizes for accurate and precise estimation (De Ayala, 2009; Kirisci et al., 2001; Reise & Yu, 1990). At this point, it is important to explain what accuracy and precision are.

Accuracy indicates how close the measured values are to known values. For example, if in the laboratory one measures a given object as 132.2 cm, but the known height is 150 cm, then the measurement of the given object is inaccurate. In this case, the measurement is not close to the known value. Precision indicates how two or more measurements are close to each other. Using the aforementioned example, if one measures a given object ten times, and obtains 132.2 cm each time, then the measurement of that object is very precise. Any measurement can be very precise but inaccurate, as described above, while it can also be accurate but imprecise (Barış Pekmezci & Gülleroğlu, 2019).

Sample sizes and test lengths are the other independent variables of this research besides number of replications and IRT models. In order to determine which sample sizes and test lengths were commonly used in unidimensional IRT studies, literature was reviewed. The literature review results are given in Table 2.

Table 2. *Literature review about sample sizes and test lengths for unidimensional IRT models.*

Studies	Sample Size			Test Lengths
	1PL model	2PL model	3PL model	
Lord, 1968	1000	-	-	50
Hulin et al., 1982	-	-	500/1000	30/ 60
Thissen & Wainer, 1982	1000	2500	-	
Goldman & Raju, 1986	250	1000	-	
Yen, 1987	-	-	1000	10/ 20/40
Patsula & Gessaroli, 1995	-	-	1000	20/40
Baker, 1998	-	500	-	50
De La Torre & Patz, 2005	-	-	1000	10/30/ 50
Gao & Chen, 2005	-	-	500/ 2000	10/ 30/ 60
Yang, 2007	100/500/1000	-	-	15/ 30/ 45
Babcock, 2011	-	1000/2500/4000	-	54/62/70
Chuah et al., 2006	-	-	500/1000	20
Sahin & Anil, 2017	150/ 250/ 350/ 500/ 750/ 1000/2000/ 3000/ 5000	150/ 250/ 350/ 500/ 750/ 1000/2000/ 3000/ 5000	150/ 250/ 350/ 500/ 750/ 1000/2000/ 3000/ 5000	10/20/30
Matlock Cole & Paek, 2017	-	1500	3000	20/40
Ames et al., 2020	-	250/500/1000	250/500/1000	10/40

Sample sizes and test lengths differ as can be seen from Table 2. Accordingly, sample sizes are varied between 150 and 5000, while test lengths are varied between 10 and 70. Minimum sample size was determined as 500, medium sample sizes were determined as 1000 and 2000, and maximum sample size was determined as 3000 for this research. Test lengths were selected as 25 items for short tests and 50 items for long tests for this research.

To begin the simulation, the item difficulty parameters (b), the item discrimination parameters (a), the item lower asymptote parameters or guess parameters (c), and the ability parameters (θ) were chosen according to the literature review. In this study, the b parameters are normally distributed [b~N(0.50, 1.50)]; the a parameters are uniformly distributed [a~U(1.5, 2.0)], the c parameters are beta distributed [c~$Beta$(20, 90)], and the ability parameters (θ) are normally

distributed [$\theta \sim N(0, 1)$] (Bahry, 2012; Bulut & Sünbül, 2017; Cohen et al., 1993; DeMars, 2002; Feinberg & Rubright, 2016; Jiang et al., 2016; Harwell & Baker, 1991; Mislevy & Stocking, 1989; Mooney, 1997). According to these parameters, dichotomous response patterns were generated for selected conditions (3x2x4x8), which are shown in Table 3. The generation of the data sets in the test conditions, determined in the research, by two computers with 2.7 GHz Intel Core i5 8 GB RAM and 1.8 GHz Turbo Intel Core i7 16 GB RAM took approximately a month.

Table 3. *Simulation conditions.*

IRT Models	Test lenghts	Sample Size	Number of Replications							
			20	78	156	312	625	1250	2500	3000
1PL model	25 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓
	50 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓
2PL model	25 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓
	50 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓
3PL model	25 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓
	50 items	500	✓	✓	✓	✓	✓	✓	✓	✓
		1000	✓	✓	✓	✓	✓	✓	✓	✓
		2000	✓	✓	✓	✓	✓	✓	✓	✓
		3000	✓	✓	✓	✓	✓	✓	✓	✓

2.2. Simulating Model Parameters and Item Responses

All parameters were simulated based on the null (ideal) model. Any departure from the null model can cause misfit or non-fit of the data, therefore; misspecified models are not in the scope of this research. To simulate dichotomous item responses and estimate the item parameters based on the unidimensional IRT models, the "itemrecovery" function, which is composite of R functions and defined by Bulut and Sünbül (2017), was revised for this study and used. This function, which generates item parameters, simulates item responses concerning parameters, estimates the item parameters of related IRT models, and computes bias statistics, was adapted to the current study. IRT model parameters and model fit values were estimated using the mirt package (Chalmers, 2012) in R. After all bias statistics had been calculated, the relevant graphics were drawn by using the lattice package (Sarkar, 2008) in R.

2.3. Estimation of Model Parameters and Type I Error Rates

The evaluation of the accuracy and precision of item and ability parameter estimations throughout the replications was carried out via mean error (ME), root mean square error (RMSE) and standard error of estimates (SE). Mean Error (ME) measures the average magnitude of the errors. ME is the average of the differences between the model's predicted and actual values, where all individual differences have equal weight. ME is given in Equation 2:

$$ME = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}_i \quad \text{(Equation 2)}$$

where N is the total test length, $\hat{y}_j$ is the estimated item parameter for item i ($i = 1, 2, \ldots, N$), and y_j is the true item parameter for item i.

RMSE is the square root of the variance of the residuals. It indicates the fit of the model, which is the closeness of the observed data points to the model's predicted values. RMSE can range from *0* to ∞ and lower values mean better fit. The errors are squared before they are averaged. RMSE should be used when undesirable large errors exist because, in the calculation, RMSE gives a relatively high weight to large errors.

RMSE is in the same unit as the response variable and can be interpreted as the variation of unexplained variance. RMSE is an important criterion of estimation accuracy, and it is important when the interest is in the model prediction. There is no one best model fit measure; researchers should choose depending on their objectives, and more than one is often useful. RMSE is given in Equation 3:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad \text{(Equation 3)}$$

Standard Error (SE), like standard deviation, is a measure of dispersion. However, while the standard deviation is a measure of dispersion from sample values, the standard error is a measure of dispersion from the sampling distribution, which belongs to the population of interest. SE is the measure of how accurate and precise the sample is. SE is not only a measure of dispersion and accuracy of the sample statistic but also an important indicator of reliability of estimation of the population parameter. SE is given in Equation 4:

$$SE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y} - \frac{\sum_{i=1}^{N}\hat{y}}{N})^2 \quad \text{(Equation 4)}$$

In addition to bias estimation of model parameters, Type I error rates for model fit were calculated in this study. Glass et al. (1972) emphasize that sampling error contaminates empirical Type I error and statistical power. Therefore, in comparing Type I error, they highly recommended taking this sampling error into account. Glass et al. (1972) suggested Equation 5 (Type I error rates) about standard error of a sampling proportion by using the number of replications as a sample size:

$$\hat{\sigma}_p = \sqrt{\frac{(1-P)P}{R}} \quad \text{(Equation 5)}$$

where R denotes the number of replications, P is the nominal or theoretical Type I error (.05 for this study), and p is the empirical or the observed Type I error. Glass et al. (1972) advise against considering the difference between a particular observed p value and the theoretical P value significant, if departure is less than two standard errors of that p.

To estimate accuracy of error rate, the MC variance of an estimate of Type I error rate ($\frac{\hat{\sigma}_p}{\sqrt{R}}$) was used, where $\hat{\sigma}_p$ is the simulated standard deviation of the p values, and R is the number of replications.

3. RESULT / FINDINGS

Findings are given in the order of the research questions. Most parts of the analysis outputs are given in the Supplementary file due to the excessive number of simulation conditions (in total 192 conditions from Table 3). Only the most remarkable findings are given via figures and tables in the findings section. For detailed information, the Supplementary file can be reviewed.

3.1. The Effect of the Number of Replications on Estimation of Item Parameters with Varying Sample Sizes and Test Lengths

Bias estimations of item parameters obtained by examining simulation conditions are given in this section. Figure 2, Figure 3, and Figure 4 summarize for RMSE values according to IRT models. Besides, ME, RMSE, and SE values are given in the Supplementary file.

When ME, RMSE, and SE values according to item parameters were examined, the same pattern was seen for all IRT models. Therefore, findings were interpreted in a way that concerns all IRT models. Increasing the sample size resulted in decreasing RMSE values for *b* parameters in all simulation conditions. When the RMSE values were examined in terms of sample sizes in detail, for all replication numbers, it was seen that the bias differences between samples were quite large. Contrary to this, when each sample was analyzed within itself, a slight difference was found in regard to replication number. For example, for the simulation condition with the 1PL model with a test length of 25 items and sample size of 1000, RMSE values obtained from 5000 replications and 78 replications were compared, the difference between them was found to be 0.001. This indicates that parameter estimation accuracy was mostly affected by sample size rather than by the number of replications. Results of ME, RMSE, and SE can be seen in Table 4.

Table 4. *Accuracy and precision of b parameters.*

IRT Models	Test lenghts	Bias statistics	Number of replications			
			20		5000	
Sample size			500	3000	500	3000
1PL model	25 items	ME	0.024	0.007	0.002	0.000
		RMSE	0.024	0.055	0.134	0.054
		SE	1.496	1.453	1.465	1.457
	50 items	ME	-0.002	-0.003	0.002	0.000
		RMSE	0.139	0.053	0.135	0.055
		SE	1.510	1.484	1.489	1.483
2PL model	25 items	ME	-0.01	-0.009	0.008	0.003
		RMSE	0.196	0.073	0.191	0.075
		SE	1.517	1.479	1.481	1.458
	50 items	ME	0.037	-0.015	0.009	0.003
		RMSE	0.199	0.085	0.190	0.075
		SE	1.529	1.479	1.509	1.482
3PL model	25 items	ME	-0.030	-0.002	-0.050	-0.004
		RMSE	0.628	0.228	0.588	0.203
		SE	1.731	1.470	1.621	1.463
	50 items	ME	-0.046	-0.005	-0.047	-0.003
		RMSE	0.553	0.172	0.519	0.185
		SE	1.668	1.450	1.602	1.489

When ME and SE statistics were examined, although the average ME and SE did not change as much as RMSE values according to the sample size, the highest bias values were observed in the smallest sample size for both test lengths. Additionally, except for the 3PL model in terms

of RMSE values, *b* parameter estimation biases were found to be quite similar for both test lengths. According to SE values, it can be said that the precision of *b* parameter estimates were not much affected by the number of replications. Accuracy and precision of *b* parameters, which was obtained with the minimum replication number (20) and the largest sample size (3000), could not be obtained with the maximum replication number (5000) and the minimum sample size (500).

For *a* parameters, bias statistics were examined and interpreted in detail according to both test lengths. In regard to *a* parameters, increasing the sample size resulted in decreased bias statistics (ME, RMSE, SE) for both test lengths except one condition. Estimation of *a* parameter accuracy and precision is directly related with sample size. Accuracy and precision of *a* parameters, which was obtained with the minimum replication number (20) and the largest sample size (3000), could not be obtained with the maximum replication number (5000) and the minimum sample size (500). Related findings can be seen in Table 5.

Table 5. *Accuracy and precision of a parameters.*

IRT Models	Test lenghts	Bias statistics	Number of replications			
			20		5000	
Sample size			500	3000	500	3000
2PL model	25 items	ME	0.018	0.002	0.022	0.004
		RMSE	0.214	0.086	0.215	0.085
		SE	0.256	0.167	0.249	0.163
	50 items	ME	0.003	0.006	0.020	0.003
		RMSE	0.215	0.082	0.207	0.082
		SE	0.248	0.163	0.244	0.163
3PL model	25 items	ME	0.201	0.008	0.179	0.008
		RMSE	0.735	0.194	0.649	0.194
		SE	0.703	0.229	0.631	0.229
	50 items	ME	0.168	0.019	0.146	0.016
		RMSE	0.592	0.163	0.544	0.168
		SE	0.583	0.211	0.538	0.217

Regardless of the sample size, bias statistics (ME, RMSE, SE) were not substantially affected by the number of replications. For example, for the 2PL model with a test length of 50 items and sample size of 500, the SE of the *a* parameters obtained from 5000 replications and 20 replications were compared, and the difference between them was found to be 0.004. Regardless of the sample size, parameter estimation bias (ME, RMSE, SE) of *a* parameters were not affected by the number of replications, as in *b* parameters. In summary, it was seen that the sample size had the largest effect rather than the number of replications in the estimation of both *a* and *b* parameters in both test lengths.

For *c* parameters, bias statistics were examined, and it was seen that as the sample size increased, SE and RMSE decreased. When the sample size was the largest (3000), the estimation accuracy and precision obtained with the minimum replication number (20) could not be obtained with the smallest sample size (500) and maximum replication number (5000). Related findings can be seen in Table 6. In summary, like the other item parameters (*a* and *b*), sample size had a greater effect on *c* parameter estimation bias than replication number.

When the effect of test lengths on parameter estimation bias was examined, it was seen that, for *a* parameters, increasing the length of the test provided more accurate and precise parameter estimation in all sample sizes and replication numbers. Increasing the length of the test

decreased the estimation bias of *a* parameters. For *b* parameters, increasing the test lengths resulted in increased SE. In terms of RMSE values, there was no remarkable change in the accuracy of *b* parameter estimations. In general, increasing the test lengths resulted in increased accuracy and precision of *c* parameters.

Table 6. *Accuracy and precision of c parameters.*

IRT Model	Test lenghts	Bias statistics	Number of replications			
			20		5000	
Sample size			500	3000	500	3000
3PL model	25 items	ME	-0.001	-0.007	0.002	-0.002
		RMSE	0.134	0.076	0.141	0.077
		SE	0.144	0.082	0.141	0.083
	50 items	ME	0.001	-0.004	0.004	-0.002
		RMSE	0.134	0.072	0.131	0.071
		SE	0.134	0.078	0.133	0.078

Figure 2. *RMSE values for 1PL model.*

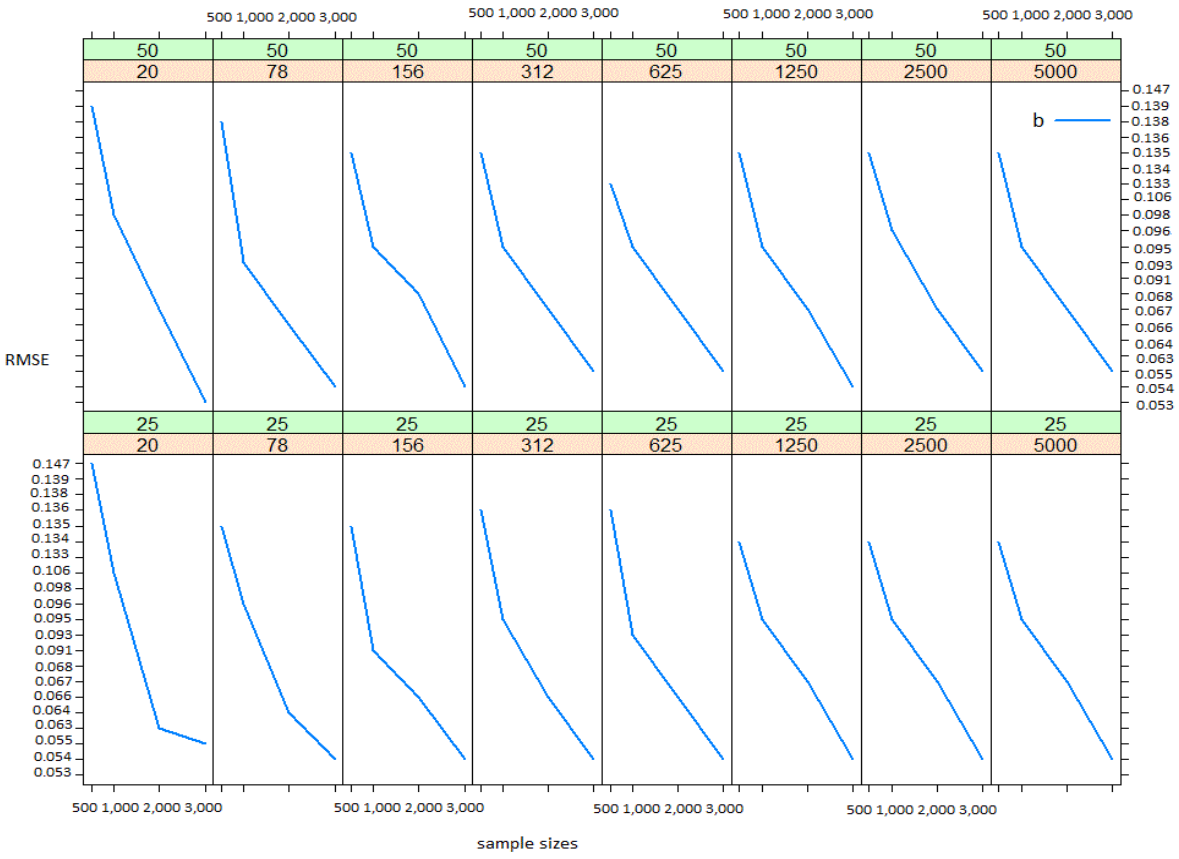


Figure 3. *RMSE values for 2PL model.*

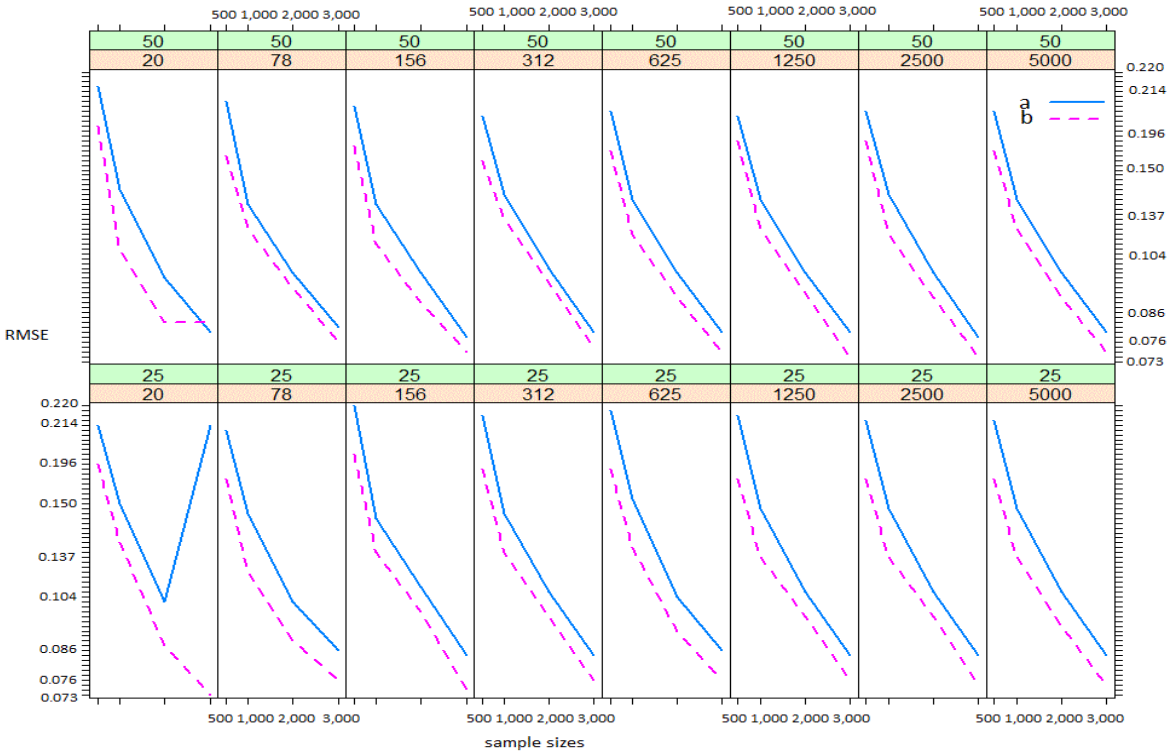
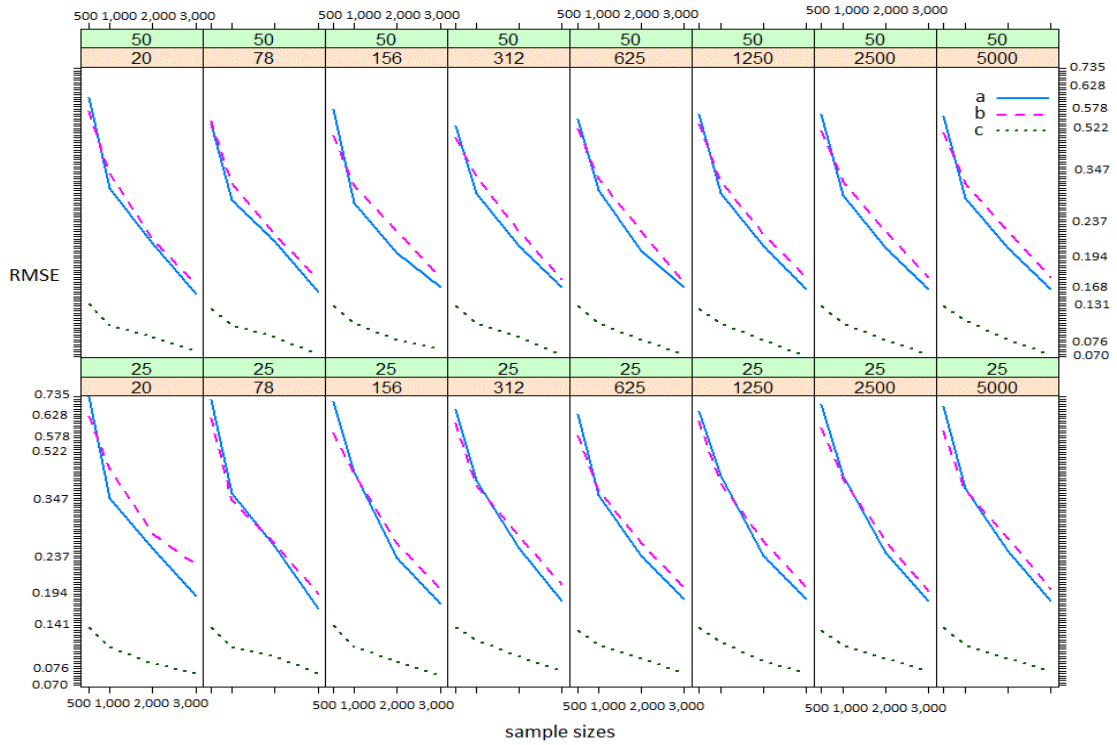


Figure 4. *RMSE values for 3PL model.*



As a result, it was seen that the sample size had a greater effect rather than the number of replications in the estimation of item parameters (*a*, *b*, and *c*). The parameter accuracy and precision obtained with the minimum replication number when the sample size was the largest could not be obtained with the maximum replication number when the sample size was the

smallest. When item parameter biases were examined among IRT models, the largest bias values were obtained from the 3PL model. It can be concluded that the increase in the number of parameters estimated by the model resulted in more biased estimates.

3.2. The Effect of the Number of Replications on Estimation of Ability Parameters with Varying Sample Sizes and Test Lengths

Bias estimations of ability parameters (θ) obtained by examining simulation conditions are given in this section. Besides, all bias statistics are given in the Supplementary file. The ability parameter (θ) estimation accuracy and precision did not change much according to test lengths within all IRT models. Apart from this finding, between all IRT models, some minor differences occurred in terms of bias statistics.

For the 1PL model, when the bias statistics were inspected in detail, it was seen that in general, estimation accuracy for θ parameters increased if sample size was increased. When SE values were investigated in terms of estimation precision, the largest sample size (3000) and minimum replication number (20) conditions (0.071 and 0.044, respectively for test lengths 25 and 50 items) were superior to the smallest sample size (500) and maximum replication number (5000) conditions (0.176 and 0.087, respectively for test lengths 25 and 50 items). In other words, θ parameters with the minimum sample size and maximum replication number were not predicted as accurately as with the large sample size and minimum replication number. Increasing the sample size would provide more precise θ parameters. Lastly, when the RMSE values for θ parameters were analyzed, it can be said that the accuracy of θ parameters increased as the sample size was increased.

For the 2PL model, when the RMSE values regarding θ parameters were examined, it can be said that the accuracy of θ parameter estimations increased as sample size was increased for both test lengths. When the SE statistics were analyzed, it was detected that θ parameters were estimated most precisely in the 2000-sample size for both test lengths. When the effects of test length in the estimation of θ parameters were examined, there were not seen many differences in terms of bias statistics.

For the 3PL model, the estimation accuracy of θ parameters increased with increasing sample size for both test lengths. In general, regardless of the sample size, the number of replications did not have a remarkable effect on the accuracy and precision of θ parameters. However, the number of replications did have an important effect on the precision of θ parameter estimations when for the test length of 50 items and the sample size was 1000. According to findings the sample size had a greater effect on the estimation accuracy of θ parameters than the number of replications for all IRT models.

3.3. The Effect of the Number of Replications on Estimation of Model Fit with Varying Sample Sizes and Test Lengths

Model fit statistics (M_2 and $RMSEA_2$) were evaluated for all IRT models. M_2 and $RMSEA_2$ statistics are given respectively in Figure 5 and Figure 6 for all IRT models. According to M_2 values, increasing the test length did not show improvement on the model fit. Additionally, when $RMSEA_2$ values were examined for the 1PL model, the best model fit was seen in the largest sample size for both test lengths. For both the 2PL model and 3PL model, increasing the test length resulted in decreased/poor model fit in terms of M_2 values. Although not much change was seen, $RMSEA_2$ values decreased to some extent regardless of the sample size for both the 2PL model and 3PL model. Lastly, it was also detected that regardless of the sample size, the number of replications had no effect on model fit values for both test lengths for all IRT models.

Figure 5. *M_2 values for all IRT models.*

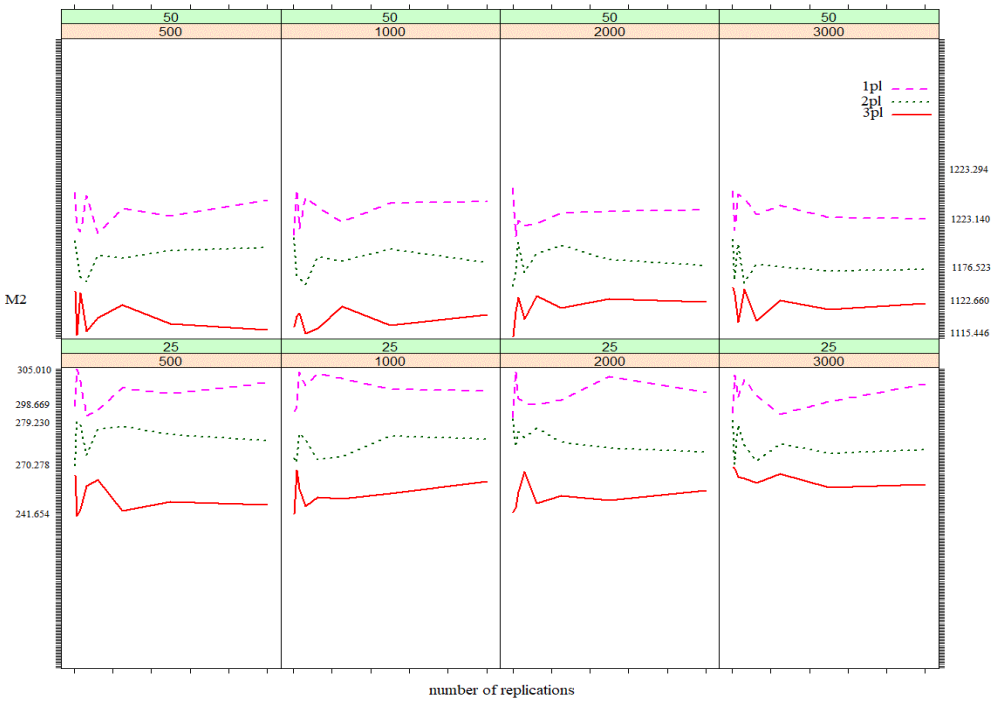
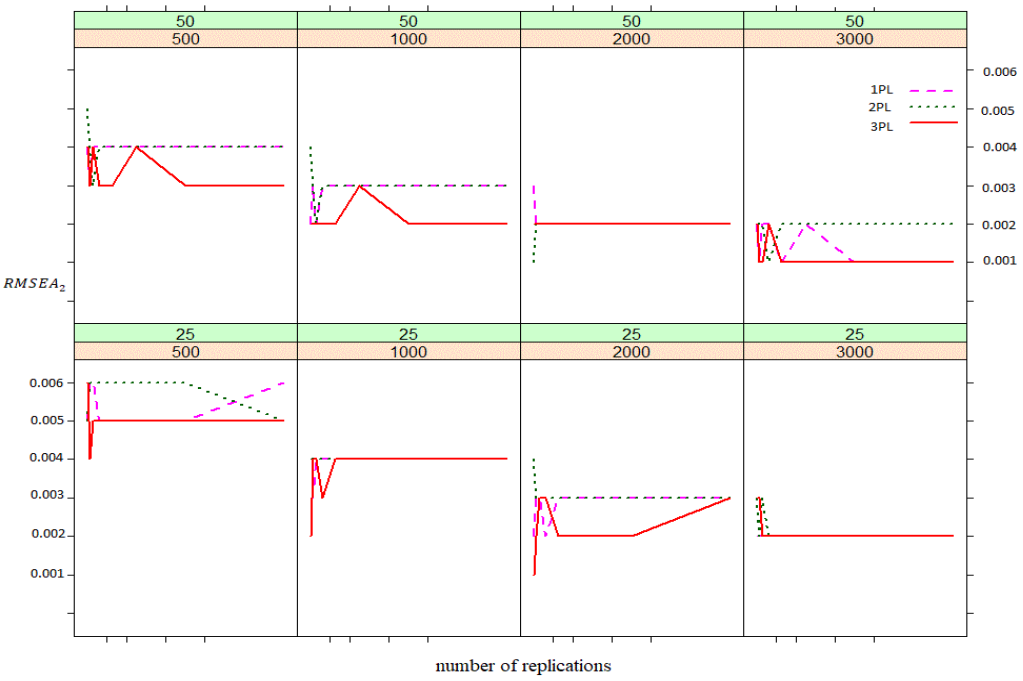


Figure 6. *$RMSEA_2$ values for all IRT models.*



In addition to these findings, Type I error inflation rates were calculated according to Glass et al. (1972), and these are presented in Appendix A, Appendix B, and Appendix C. The difference between a particular empirical alpha (p) value and the nominal alpha (P) value was indicated as significant if departure was two standard errors of p. When Type I error inflation rates are examined in Appendix A, it is seen that Type I error inflation was only seen at 20 replications for the 1PL model in all sample sizes and test lengths. Also, when test length of 50 items, 78 replications were enough for actual model fit interpretations for the sample sizes 500 and 1000.

For the 2PL model, Type I error inflation, given in Appendix B, was only seen at 20 replications for the test length of 25 items in all sample sizes except when the sample size was 3000. When the sample size was 3000, Type I error inflation was seen at 78 replications also. Type I error inflation was seen at only 20 replications in all sample sizes for the test length of 50 items.

For the 3PL model, in all sample sizes and test lengths Type I error inflation was seen at 20 replications. Additionally, Type I eror inflations, given in Appendix C, were seen at 78 replications in both 500 and 3000 sample sizes for the test length 25 items. For the test length 50, Type I error inflation was 78 replications in only 2000 sample sizes. In summary, Type I error rates were not affected except at 20 and 78 replications for all IRT models.

Accuracy of error rate estimation and confidence intervals of empirical alpha (p), given in Appendices (A, B, and C), were examined and the same results were achieved for all IRT models. It is important to underline the finding that accuracy of error rate estimation did not change according to either test length or sample size and was affected more by the replication number. The lowest accuracy of error rate was seen at 20 replications for the 1PL model, and at 20 and 78 replications for both the 2PL model and 3PL model. Lastly, the largest confidence interval of empirical alpha (p) was seen in the smallest replication number, and that is important in terms of supporting inferences about accuracy.

The main concern of this study is determining a suitable replication number for simulations different test conditions. When test conditions which are determined in this research are considered, findings show that the number of replication effects Type I error inflation. Type I error inflation was seen at 20 and 78 replications. In general, it can be thought that 156, 312, or 625 replications may enough for avoiding Type I error inflation (see Supplementary file for details). However, other factors, such as item parameter estimation and model fit considered together, it is suggested that at least 625 replications should be performed in terms of Type I error rates.

4. DISCUSSION and CONCLUSION

The purpose of the current study was to determine the required number of replications for the most accurate and precise parameter estimations in conducting MC simulation studies involving unidimensional IRT models. In line with research purpose, different sample sizes and different test lengths were defined as test conditions besides the number of replications.

The first major finding was that neither the test length nor the replication numbers had an effect on item parameter estimation accuracy and precision for all IRT models. On the contrary, the sample size had the largest effect rather than the number of replications in estimation of item parameters. It can be concluded that when the sample is large, even with the smallest number of replications, item parameters can be estimated with adequate precision and accuracy.

Consistent with the current research, Hulin et al. (1982) showed that in the studies of item bias which place emphasis on accuracy, large numbers of items were not necessarily needed. However, they recommended using large samples to obtain accurate item parameter estimates. Besides, they proved that a sample size of 500 for the 2PL model and 1000 for the 3PL model was needed, but also underlined that the more accurate results appeared with a sample size of 2000. Also, consistent with the present study, Ames et al. (2020) found that difficulty parameters had smaller mean bias as sample size was increased for the 2PL model. However, contrary to the present study, they found that increasing the sample size increased the mean bias of discrimination parameters.

When item parameter biases were examined among IRT models, the largest bias values were obtained from 3PL model. It can be concluded that the increase in the number of parameters estimated by the model resulted in more biased estimates.

The study also showed that the best way to increase estimation accuracy of θ parameters was to increase the sample size. Contrary to this, θ parameters were most precisely estimated among other samples only with 2000 for the 2PL model and 3000 for the 3PL model, and increased test length had no effect on estimation precision like the 1PL model. For the 2PL model and 3PL model, only sample size had an effect on estimation in terms of estimation accuracy of θ parameters. The largest sample size had a larger effect on estimation accuracy than the number of replications in both test lengths for all IRT models. This is also consistent with the findings of Hulin et al. (1982), who reported that ability estimates were less accurate in small sample sizes for the 3PL model.

The second major finding was that although the number of replications did not seem to have an effect on the model fit, it was decisive in Type I error inflation and error prediction accuracy for all IRT models. Besides, the most determining factor in model fit was the sample size and long tests had relatively better fit values than short tests. This finding is consistent with that of Schumacker et al. (1994), who found no differences between Rasch item and ability fit statistics based on the number of replications, and the Type I error rates were close to expected values. In accordance with the present study, they recommended being more sensitive to the sample size and test length.

The most obvious finding to emerge from this study was that the sample size had the most important effect on estimation bias for both item parameters and model fit statistics. However, the number of replications was found to be effective on Type I error inflation. Generally, when the number of replications is 20 and 78, Type I error inflation was seen much as per other conditions. When all test conditions determined in this study, especially the acccuracy of error rate estimate were evaluated together, accuracy of error rate estimate was seen too close to zero for 625 replications. Besides, also 156 replications and above can be recommended but if the researchers want to get more accurate results, should perform at least 625 replications.

The present study investigated the effect of replication number on the estimation of item and ability estimations and model fit statistics in the MC method based on unidimensional IRT models. It was concluded that the number of replications was not a very impressive factor in the test conditions determined in this study for unidimensional IRT models. In particular, it is seen that sample size is the most effective factor in the estimation of the item and ability parameter and model fit. However, it was concluded that the number of replications is effective in estimating Type I error inflation and accuracy of error rate estimate. In general, as a conclusion of this study, when studying with unidimensional IRT models, it is highly recommended that researchers use large samples instead of studying with small samples and excessive replications.

This study showed that an increase in the number of parameters estimated by the model resulted in increased bias. Therefore, it should be taken into consideration that the adequate number of replications would differ in multi-dimensional models because of increasing estimations of the number of parameters. Similarly, since this study focused on IRT models used with dichotomous items, similar studies could be carried out with polytomous items. All simulations and analyses were performed according to the null (ideal) model. Further research can focus on determining the ideal replication number for misfit data. Due to the fact that it is a simulation study, it is suggested that new studies are conducted on the same condition for generalizations.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Fulya Baris-Pekmezci: Investigation, Resources, Visualization, Software, Analyze, and Writing. **Asiye Sengul-Avsar**: Investigation, Methodology, Analyze, Supervision, Validation, and Writing.

ORCID

Fulya BARIŞ PEKMEZCİ https://orcid.org/0000-0001-6989-512X
Asiye ŞENGÜL AVŞAR https://orcid.org/0000-0001-5522-2514

5. REFERENCES

Ames, A. J., Leventhal, B. C., & Ezike, N. C. (2020). Monte Carlo simulation in item response theory applications using SAS. *Measurement: Interdisciplinary Research and Perspectives*, *18*(2), 55-74. https://doi.org/10.1080/15366367.2019.1689762

Babcock, B. (2011). Estimating a noncompensatory IRT model using Metropolis within Gibbs sampling. *Applied Psychological Measurement*, *35*(4), 317-329. http://dx.doi.org/10.1177/0146621610392366

Bahry, L. M. (2012). Polytomous item response theory parameter recovery: an investigation of nonnormal distributions and small sample size [Master's Thesis]. ProQuest Dissertations and Theses Global.

Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a Gibbs sampling procedure. *Applied Psychological Measurement*, *22*(2), 153-169. https://doi.org/10.1177/01466216980222005

Baldwin, P. (2011). A strategy for developing a common metric in item response theory when parameter posterior distributions are known. *Journal of Educational Measurement, 48*(1), 1-11. Retrieved December 9, 2020, from http://www.jstor.org/stable/23018061

Barış Pekmezci, F., & Gülleroğlu, H. (2019). Investigation of the orthogonality assumption in the bifactor item response theory. *Eurasian Journal of Educational Research, 19*(79), 69-86. http://dx.doi.org/10.14689/ejer.2019.79.4

Bulut, O., & Sünbül, Ö. (2017). Monte Carlo simulation studies in item response theory with the R programming language. *Journal of Measurement and Evaluation in Education and Psychology, 8*(3), 266-287. https://doi.org/10.21031/epod.305821

Cai, L., & Thissen, D. (2014). *Modern Approaches to Parameter Estimation in Item Response Theory from: Handbook of Item Response Theory Modeling, Applications to Typical Performance Assessment.* Routledge.

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29. https://doi.org/10.18637/JSS.V048.I06

Chuah, S. C., Drasgow, F., & Luecht, R. (2006). How big is big enough? Sample size requirements for CAST item parameter estimation. *Applied Measurement in Education*, *19*(3), 241-255. https://doi.org/10.1207/s15324818ame1903_5

Cohen, A. S., Kim, S. H., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement, 17*(4), 335-350. https://doi.org/10.1177/014662169301700402

Crişan, D. R., Tendeiro, J. N., & Meijer, R. R. (2017). Investigating the practical consequences of model misfit in unidimensional IRT models. *Applied Psychological Measurement*, *41*(6), 439-455. https://doi.org/10.1177/0146621617695522

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando: Harcourt Brace Jovanovich Inc.

De Ayala, R. J. (2009). *The theory and practice of item response theory.* New York, NY: Guilford Press.

De La Torre, J., & Patz, R. J. (2005). Making the most of what we have: A practical application of multidimensional item response theory in test scoring. *Journal of Educational and Behavioral Statistics*, *30*(3), 295-311. https://www.jstor.org/stable/3701380

DeMars, C. E. (2002, April). Recovery of graded response and partial credit parameters in multilog and parscale. Annual meeting of American Educational Research Association, Chicago. https://commons.lib.jmu.edu/cgi/viewcontent.cgi?article=1034&context=gradpsych

Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, *35*(2), 36-49. https://doi.org/10.1111/emip.12111

Fu, J. (2019). *Maximum marginal likelihood estimation with an expectation–maximization algorithm for multigroup/mixture multidimensional item response theory models* (No. RR-19-35). ETS Research Report Series, https://doi.org/10.1002/ets2.12272

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*(4), 351-380. https://doi.org/10.1207/s15324818ame1804_2

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237-288. https://doi.org/10.3102/00346543042003237

Goldman, S. H., & Raju, N. S. (1986). Recovery of one-and two-parameter logistic item parameters: An empirical study. *Educational and Psychological Measurement*, *46*(1), 11-21. https://doi.org/10.1177/0013164486461002

Hair, J. F., Black W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis.* (8th edition). Annabel Ainscow.

Han, K. T. (2007). WinGen: windows software that generates irt parameters and item responses. *Applied Psychological Measurement, 31*(5), 457-459. https://doi.org/10.1177/0146621607299271

Hanson, B. A. (1998, October). IRT parameter estimation using the EM algorithm. http://www.b-a-h.com/papers/note9801.pdf

Harwell, M. (1997). Analyzing the results of monte carlo studies in item response theory. *Educational and Psychological Measurement, 57*(2), 266-279. https://doi.org/10.1177/0013164497057002006

Harwell, M. R., & Baker, F. B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: A didactic. *Applied Psychological Measurement, 15*(4), 375–389. https://doi.org/10.1177/014662169101500409

Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement, 20*(2), 101-125. https://doi.org/10.1177/014662169602000201

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement, 6*(3), 249–260. https://doi.org/10.1177/014662168200600301

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in psychology*, *7*, 109. https://doi.org/10.3389/fpsyg.2016.00109

Kirisci, L., Hsu, T. C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, *25*(2), 146-162. https://doi.org/10.1177/01466210122031975

Kleijnen, J. P. (1987). *Statistical tools for simulation practitioners*. Marcel Dekker.

Lee, S., Bulut, O., & Suh, Y. (2017). Multidimensional extension of multiple indicators multiple causes models to detect DIF. *Educational and Psychological Measurement*, *77*(4), 545–569. https://doi.org/10.1177/0013164416651116

Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, *28*(4), 989-1020. https://doi.org/10.1177/001316446802800401

Matlock, K. L., & Turner, R. (2016). Unidimensional IRT item parameter estimates across equivalent test forms with confounding specifications within dimensions. *Educational and Psychological Measurement*, *76*(2), 258-279. https://doi.org/10.1177/0013164415589756

Matlock Cole, K., & Paek, I. (2017). PROC IRT: A SAS procedure for item response theory. *Applied Psychological Measurement*, *41*(4), 311-320. https://doi.org/10.1177/0146621616685062

McDonald, R. P. (1982). Linear Versus Models in Item Response Theory. *Applied Psychological Measurement*, *6*(4), 379-396. https://doi.org/10.1177/014662168200600402

Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement*, *13*(1), 57-75. https://doi.org/10.1177/014662168901300106

Mooney, C. Z. (1997). *Monte Carlo simulation*. Thousand Oaks, CA: Sage.

Mundform, D. J., Schaffer, J., Kim, M. J., Shaw, D., Thongteeraparp, A., & Supawan, P. (2011). Number of replications required in Monte Carlo simulation studies: A synthesis of four studies. *Journal of Modern Applied Statistical Methods*, *10*(1), 19-28. https://doi.org/10.22237/jmasm/1304222580

Park, Y. S., Lee, Y. S., & Xing, K. (2016). Investigating the impact of item parameter drift for item response theory models with mixture distributions. *Frontiers in Psychology*, *7*, 255. https://doi.org/10.3389/fpsyg.2016.00255

Patsias, K., Sheng, Y., & Rahimi, S. (2009, September 24-26). A high performance Gibbs sampling algorithm for item response theory. 22nd International Conference on Parallel and Distributed Computing and Communication Systems, Kentucky, USA.

Patsula, L. N., & Gessaroli, M. E. A (1995, April). Comparison of item parameter estimates and iccs produced. https://files.eric.ed.gov/fulltext/ED414333.pdf

Preecha, C. (2004). Numbers of replications required in ANOVA simulation studies [Doctoral dissertation, University of Northern Colorado]. ProQuest Dissertations and Theses Global.

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, *27*(2), 133-144. https://www.jstor.org/stable/1434973

Reise, S., Moore, T., & Maydeu-Olivares, A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models. *Educational and Psychological Measurement*, *71*(4), 684-711. https://doi.org/10.1177/0013164410378690

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement*, *26*(2), 192-207. https://doi.org/10.1177/01421602026002006

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method.* John Wiley and Sons, New York. https://doi.org/10.1002/9780470316511

Sahin, A., & Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17(1)*, 321-33. https://doi.org/10.12738/estp.2017.1.0270

Sarkar, D. (2008). *Lattice: multivariate data visualization with R*. Springer, New York.

Schumacker, R. E, Smith, R. M., & Bush, J. M. (1994, April). *Examining replication effects in Rasch fit statistics.* American Educational Research Association Annual Meeting, New Orleans.

Sen, S., Cohen, A. S., & Kim, S. H. (2016). The impact of non-normality on extraction of spurious latent classes in mixture IRT models. *Applied Psychological Measurement*, *40*(2), 98-113. https://doi.org/10.1177/0146621615605080

Sheng, Y., & Wikle, C. K. (2007). Comparing multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, *67*(6), 899-919. https://doi.org/10.1177/0013164406296977

Tavares, H. R., Andrade, D. F. D., & Pereira, C. A. D. B. (2004). Detection of determinant genes and diagnostic via item response theory. *Genetics and Molecular Biology*, *27*(4), 679-685. https://doi.org/10.1590/S1415-47572004000400033

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, *47*(4), 397-412. https://doi.org/10.1007/BF02293705

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Amer Psychological Assn.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. Guilford Press.

Thompson, N. A. (2009). Ability estimation with item response theory. *Assessment Systems Corporation.* https://assess.com/docs/Thompson_(2009)_-_Ability_estimation_with_IRT.pdf

Şengül Avşar, A., & Tavşancıl, E. (2017). Examination of polytomous items' psychometric properties according to nonparametric item response theory models in different test conditions. *Educational Sciences: Theory & Practice, 17*(2). https://doi.org/10.12738/estp.2017.2.0246

van der Linden, W. J. (Ed.). (2018). *Handbook of item response theory, three volume set*. CRC Press.

van Onna, M. J. H. (2004). Ordered latent class models in nonparametric item response theory. [Doctoral dissertation]. University of Groningen.

Weissman, A. (2013). Optimizing information using the EM algorithm in item response theory. *Annals of Operations Research*, *206*(1), 627-646. https://doi.org/10.1007/s10479-012-1204-4

Yang, S. (2007). A comparison of unidimensional and multidimensional RASCH models using parameter estimates and fit indices when assumption of unidimensionality is violated [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations and Theses Global.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, *52*(2), 275-291. https://doi.org/10.1007/BF02294241

Zhang, B. (2008). Application of unidimensional item response models to tests with items sensitive to secondary dimensions. *The Journal of Experimental Education, 77*(2), 147-166. https://doi.org/10.3200/JEXE.77.2.147-166

6. APPENDIX

6.1. Appendix A

Table A1. *Type I error rate and accuracy of error estimate from 25 items for 1PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.493	0.443	0.486	0.499	0.000
	2500	0.502	0.452	0.493	0.511	0.000
	1250	0.515	0.465	0.503	0.527	0.000
	625	0.495	0.445	0.478	0.512	0.001
	312	0.488	0.438	0.463	0.512	0.002
	156	0.493	0.443	0.459	0.528	0.003
	78	0.485	0.435	0.436	0.535	0.006
	20	0.515	0.465	0.418	0.613	0.025
2000	5000	0.497	0.447	0.491	0.503	0.000
	2500	0.489	0.439	0.480	0.497	0.000
	1250	0.501	0.451	0.489	0.514	0.000
	625	0.501	0.451	0.483	0.518	0.001
	312	0.502	0.452	0.478	0.527	0.002
	156	0.497	0.447	0.462	0.532	0.003
	78	0.464	0.414	0.415	0.513	0.006
	20	0.521	0.471	0.424	0.619	0.025
1000	5000	0.496	0.446	0.490	0.502	0.000
	2500	0.495	0.445	0.486	0.504	0.000
	1250	0.499	0.449	0.486	0.511	0.000
	625	0.488	0.438	0.476	0.501	0.000
	312	0.481	0.431	0.464	0.499	0.001
	156	0.489	0.439	0.465	0.514	0.002
	78	0.470	0.420	0.435	0.505	0.003
	20	0.508	0.458	0.458	0.557	0.006
500	5000	0.491	0.441	0.484	0.497	0.000
	2500	0.495	0.445	0.486	0.504	0.000
	1250	0.491	0.441	0.478	0.503	0.000
	625	0.507	0.457	0.490	0.525	0.001
	312	0.527	0.477	0.502	0.552	0.002
	156	0.490	0.440	0.455	0.525	0.003
	78	0.434	0.384	0.385	0.484	0.006
	20	0.508	0.458	0.411	0.605	0.025

Table A2. *Type I error rate and accuracy of error estimate from 50 items for 1PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.505	0.455	0.498	0.511	0.000
	2500	0.504	0.454	0.496	0.513	0.000
	1250	0.499	0.449	0.487	0.511	0.000
	625	0.501	0.451	0.483	0.518	0.001
	312	0.480	0.430	0.455	0.505	0.002
	156	0.464	0.414	0.429	0.499	0.003
	78	0.526	0.476	0.477	0.575	0.006
	20	0.464	0.414	0.367	0.562	0.025
2000	5000	0.502	0.452	0.496	0.508	0.000
	2500	0.501	0.451	0.492	0.510	0.000
	1250	0.500	0.450	0.488	0.513	0.000
	625	0.507	0.457	0.490	0.524	0.001
	312	0.506	0.456	0.481	0.530	0.002
	156	0.510	0.460	0.475	0.545	0.003
	78	0.558	0.508	0.509	0.608	0.006
	20	0.411	0.361	0.313	0.508	0.025
1000	5000	0.491	0.441	0.484	0.497	0.000
	2500	0.498	0.448	0.489	0.506	0.000
	1250	0.505	0.455	0.492	0.517	0.000
	625	0.499	0.449	0.482	0.517	0.001
	312	0.477	0.427	0.452	0.502	0.002
	156	0.523	0.473	0.488	0.558	0.003
	78	0.426	0.376	0.376	0.475	0.006
	20	0.539	0.489	0.441	0.636	0.025
500	5000	0.486	0.436	0.480	0.492	0.000
	2500	0.483	0.433	0.474	0.491	0.000
	1250	0.489	0.439	0.476	0.501	0.000
	625	0.489	0.439	0.471	0.506	0.001
	312	0.487	0.437	0.462	0.511	0.002
	156	0.491	0.441	0.456	0.526	0.003
	78	0.516	0.466	0.466	0.565	0.006
	20	0.445	0.395	0.348	0.543	0.025

6.2. Appendix B

Table B1. *Type I error rate and accuracy of error estimate from 25 items for 2PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.496	0.446	0.490	0.502	0.000
	2500	0.500	0.450	0.491	0.508	0.000
	1250	0.493	0.443	0.481	0.506	0.000
	625	0.519	0.469	0.502	0.536	0.001
	312	0.496	0.446	0.472	0.521	0.002
	156	0.476	0.426	0.441	0.511	0.003
	78	0.528	0.478	0.478	0.577	0.006
	20	0.453	0.403	0.356	0.550	0.025
2000	5000	0.499	0.449	0.493	0.505	0.000
	2500	0.492	0.442	0.483	0.501	0.000
	1250	0.490	0.440	0.478	0.502	0.000
	625	0.475	0.425	0.457	0.492	0.001
	312	0.489	0.439	0.465	0.514	0.002
	156	0.483	0.433	0.448	0.518	0.003
	78	0.501	0.451	0.452	0.551	0.006
	20	0.325	0.275	0.228	0.423	0.023
1000	5000	0.489	0.439	0.483	0.495	0.000
	2500	0.486	0.436	0.477	0.494	0.000
	1250	0.502	0.452	0.490	0.515	0.000
	625	0.511	0.461	0.494	0.529	0.001
	312	0.490	0.440	0.465	0.514	0.002
	156	0.487	0.437	0.452	0.522	0.003
	78	0.509	0.459	0.460	0.558	0.006
	20	0.502	0.452	0.405	0.599	0.025
500	5000	0.491	0.441	0.485	0.498	0.000
	2500	0.486	0.436	0.477	0.495	0.000
	1250	0.476	0.426	0.463	0.488	0.000
	625	0.477	0.427	0.459	0.494	0.001
	312	0.496	0.446	0.472	0.521	0.002
	156	0.452	0.402	0.417	0.487	0.003
	78	0.451	0.401	0.402	0.500	0.006
	20	0.545	0.495	0.447	0.642	0.025

Table B2. *Type I error rate and accuracy of error estimate from 50 items for 2PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.490	0.440	0.483	0.497	0.000
	2500	0.491	0.441	0.482	0.500	0.000
	1250	0.485	0.435	0.473	0.497	0.000
	625	0.483	0.433	0.466	0.500	0.001
	312	0.515	0.465	0.490	0.540	0.002
	156	0.455	0.405	0.420	0.490	0.003
	78	0.513	0.463	0.464	0.562	0.006
	20	0.419	0.369	0.322	0.516	0.025
2000	5000	0.483	0.433	0.477	0.489	0.000
	2500	0.478	0.428	0.469	0.487	0.000
	1250	0.467	0.417	0.454	0.479	0.000
	625	0.476	0.426	0.459	0.494	0.001
	312	0.509	0.459	0.484	0.533	0.002
	156	0.428	0.378	0.393	0.463	0.003
	78	0.492	0.442	0.443	0.541	0.006
	20	0.531	0.481	0.433	0.628	0.025
1000	5000	0.481	0.431	0.475	0.487	0.000
	2500	0.470	0.420	0.461	0.479	0.000
	1250	0.480	0.430	0.468	0.492	0.000
	625	0.474	0.424	0.457	0.491	0.001
	312	0.522	0.472	0.497	0.547	0.002
	156	0.512	0.462	0.477	0.547	0.003
	78	0.510	0.460	0.461	0.559	0.006
	20	0.379	0.329	0.282	0.476	0.024
500	5000	0.471	0.421	0.465	0.478	0.000
	2500	0.471	0.421	0.463	0.480	0.000
	1250	0.478	0.428	0.466	0.490	0.000
	625	0.477	0.427	0.460	0.495	0.001
	312	0.516	0.466	0.492	0.541	0.002
	156	0.513	0.463	0.478	0.548	0.003
	78	0.476	0.426	0.427	0.525	0.006
	20	0.426	0.376	0.329	0.524	0.025

6.3. Appendix C

Table C1. *Type I error rate and accuracy of error estimate from 25 items for 3PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.517	0.467	0.510	0.523	0.000
	2500	0.519	0.469	0.510	0.527	0.000
	1250	0.517	0.467	0.504	0.529	0.000
	625	0.527	0.477	0.510	0.544	0.001
	312	0.491	0.441	0.466	0.515	0.002
	156	0.530	0.480	0.495	0.565	0.003
	78	0.498	0.448	0.448	0.547	0.006
	20	0.445	0.395	0.348	0.543	0.025
2000	5000	0.514	0.464	0.508	0.520	0.000
	2500	0.510	0.460	0.501	0.519	0.000
	1250	0.518	0.468	0.505	0.530	0.000
	625	0.500	0.450	0.483	0.518	0.001
	312	0.533	0.483	0.508	0.558	0.002
	156	0.511	0.461	0.476	0.546	0.003
	78	0.499	0.449	0.450	0.549	0.006
	20	0.564	0.514	0.467	0.662	0.025
1000	5000	0.523	0.473	0.517	0.530	0.000
	2500	0.530	0.480	0.521	0.539	0.000
	1250	0.515	0.465	0.502	0.527	0.000
	625	0.531	0.481	0.513	0.548	0.001
	312	0.543	0.493	0.518	0.567	0.002
	156	0.525	0.475	0.490	0.560	0.003
	78	0.518	0.468	0.469	0.568	0.006
	20	0.531	0.481	0.434	0.629	0.025
500	5000	0.531	0.481	0.524	0.537	0.000
	2500	0.526	0.476	0.517	0.535	0.000
	1250	0.519	0.469	0.507	0.532	0.000
	625	0.521	0.471	0.503	0.538	0.001
	312	0.539	0.489	0.515	0.564	0.002
	156	0.501	0.451	0.466	0.536	0.003
	78	0.552	0.502	0.502	0.601	0.006
	20	0.467	0.417	0.369	0.564	0.025

Table C2. *Type I error rate and accuracy of error estimate from 50 items for 3PL model.*

Sample size	Number of Replication	Empirical alpha (p)	Empirical alpha (p)-nominal (P) alpha	p-$2\hat{\sigma}_p$	p+$2\hat{\sigma}_p$	Accuracy of error rate estimate
3000	5000	0.512	0.462	0.506	0.518	0.000
	2500	0.516	0.466	0.507	0.524	0.000
	1250	0.495	0.445	0.482	0.507	0.000
	625	0.510	0.460	0.492	0.527	0.001
	312	0.505	0.455	0.480	0.530	0.002
	156	0.494	0.444	0.459	0.529	0.003
	78	0.455	0.405	0.406	0.504	0.006
	20	0.421	0.371	0.324	0.519	0.025
2000	5000	0.515	0.465	0.509	0.521	0.000
	2500	0.521	0.471	0.512	0.530	0.000
	1250	0.521	0.471	0.509	0.534	0.000
	625	0.521	0.471	0.503	0.538	0.001
	312	0.493	0.443	0.469	0.518	0.002
	156	0.522	0.472	0.487	0.557	0.003
	78	0.549	0.499	0.499	0.598	0.006
	20	0.593	0.543	0.496	0.691	0.025
1000	5000	0.512	0.462	0.505	0.518	0.000
	2500	0.520	0.470	0.511	0.529	0.000
	1250	0.522	0.472	0.510	0.535	0.000
	625	0.519	0.469	0.501	0.536	0.001
	312	0.528	0.478	0.503	0.552	0.002
	156	0.507	0.457	0.472	0.541	0.003
	78	0.498	0.448	0.449	0.548	0.006
	20	0.575	0.525	0.477	0.672	0.025
500	5000	0.526	0.476	0.520	0.532	0.000
	2500	0.520	0.470	0.511	0.528	0.000
	1250	0.541	0.491	0.528	0.553	0.000
	625	0.508	0.458	0.490	0.525	0.001
	312	0.513	0.463	0.488	0.537	0.002
	156	0.542	0.492	0.507	0.577	0.003
	78	0.592	0.542	0.543	0.642	0.006
	20	0.501	0.451	0.403	0.598	0.025

7. SUPPLEMENTARY FILE

7. 1. Supplementary File for 1PL Model

# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	3000	25	0.000	1.457	0.054	0.000	0.055	0.014	299.673	0.493	0.002	-38193.147	0.003	0.443	0.486	0.499	0.000
2500	3000	25	0.000	1.456	0.054	0.000	0.056	0.015	298.717	0.502	0.002	-38237.706	0.004	0.452	0.493	0.511	0.000
1250	3000	25	0.000	1.453	0.054	0.000	0.055	0.015	297.623	0.515	0.002	-38204.417	0.006	0.465	0.503	0.527	0.000
625	3000	25	-0.002	1.460	0.054	0.001	0.053	0.015	299.390	0.495	0.002	-38236.766	0.009	0.445	0.478	0.512	0.000
312	3000	25	-0.001	1.492	0.054	0.001	0.055	0.015	299.914	0.488	0.002	-37967.818	0.012	0.438	0.463	0.512	0.001
156	3000	25	-0.002	1.449	0.054	0.002	0.049	0.016	299.200	0.493	0.002	-38197.122	0.017	0.443	0.459	0.528	0.001
78	3000	25	0.006	1.439	0.054	-0.004	0.061	0.014	300.422	0.485	0.002	-38496.120	0.025	0.435	0.436	0.535	0.003
20	3000	25	0.007	1.453	0.055	-0.002	0.071	0.021	297.788	0.515	0.002	-38182.517	0.049	0.465	0.418	0.613	0.011
# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	25	0.001	1.456	0.067	0.000	0.062	0.018	299.396	0.497	0.003	-25477.162	0.003	0.447	0.491	0.503	0.000
2500	2000	25	0.000	1.461	0.067	0.000	0.061	0.018	300.012	0.489	0.003	-25457.807	0.004	0.439	0.480	0.497	0.000
1250	2000	25	0.001	1.458	0.067	0.000	0.063	0.018	298.774	0.501	0.003	-25464.042	0.006	0.451	0.489	0.514	0.000
625	2000	25	0.000	1.442	0.066	0.000	0.061	0.017	298.672	0.501	0.003	-25528.717	0.009	0.451	0.483	0.518	0.000
312	2000	25	-0.001	1.446	0.066	0.002	0.057	0.017	298.688	0.502	0.002	-25480.706	0.012	0.452	0.478	0.527	0.001
156	2000	25	-0.001	1.457	0.066	0.002	0.054	0.016	299.054	0.497	0.003	-25528.210	0.017	0.447	0.462	0.532	0.001
78	2000	25	-0.001	1.445	0.064	-0.002	0.060	0.015	304.377	0.464	0.003	-25325.638	0.025	0.414	0.415	0.513	0.003
20	2000	25	-0.001	1.486	0.063	0.002	0.047	0.014	296.433	0.521	0.002	-25390.219	0.049	0.471	0.424	0.619	0.011
# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	25	0.001	1.461	0.095	0.000	0.073	0.025	299.406	0.496	0.004	-12730.860	0.003	0.446	0.490	0.502	0.000
2500	1000	25	0.001	1.466	0.095	0.000	0.074	0.026	299.449	0.495	0.004	-12711.550	0.004	0.445	0.486	0.504	0.000
1250	1000	25	0.003	1.461	0.095	-0.002	0.078	0.025	299.915	0.488	0.004	-12700.168	0.006	0.438	0.476	0.501	0.000
625	1000	25	0.001	1.456	0.093	0.000	0.073	0.025	300.639	0.481	0.004	-12739.327	0.009	0.431	0.464	0.499	0.000
312	1000	25	0.000	1.481	0.095	0.002	0.065	0.024	299.650	0.489	0.004	-12687.959	0.012	0.439	0.465	0.514	0.001
156	1000	25	-0.002	1.437	0.091	0.002	0.071	0.026	301.334	0.470	0.004	-12823.181	0.017	0.420	0.435	0.505	0.001
78	1000	25	0.003	1.501	0.096	-0.001	0.078	0.027	298.567	0.508	0.003	-12605.717	0.025	0.458	0.458	0.557	0.003
20	1000	25	0.010	1.462	0.106	-0.010	0.116	0.035	297.865	0.521	0.003	-12561.626	0.049	0.471	0.424	0.619	0.011
# of replication	sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	25	0.002	1.465	0.134	0.001	0.176	0.036	299.756	0.491	0.005	-6355.279	0.003	0.441	0.484	0.497	0.000
2500	500	25	0.001	1.463	0.134	0.001	0.175	0.036	299.393	0.495	0.005	-6366.520	0.004	0.445	0.486	0.504	0.000
1250	500	25	0.003	1.460	0.134	0.001	0.173	0.035	299.465	0.491	0.005	-6363.194	0.006	0.441	0.478	0.503	0.000
625	500	25	0.001	1.460	0.136	0.002	0.176	0.036	298.152	0.507	0.005	-6360.802	0.009	0.457	0.490	0.524	0.000
312	500	25	0.002	1.441	0.136	0.001	0.183	0.037	297.027	0.527	0.004	-6390.473	0.012	0.477	0.502	0.552	0.001
156	500	25	-0.003	1.491	0.135	0.002	0.174	0.036	299.843	0.490	0.006	-6330.255	0.017	0.440	0.455	0.525	0.001
78	500	25	-0.006	1.491	0.135	0.007	0.143	0.032	305.010	0.434	0.006	-6306.223	0.025	0.384	0.385	0.484	0.003
20	500	25	0.024	1.496	0.147	-0.014	0.228	0.043	298.669	0.508	0.005	-6289.242	0.049	0.458	0.411	0.605	0.011

# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	3000	50	0.000	1.483	0.055	0.000	0.056	0.015	1223.140	0.505	0.001	-74852.745	0.003	0.455	0.498	0.511	0.000
2500	3000	50	0.000	1.480	0.055	0.000	0.056	0.014	1223.545	0.504	0.001	-74894.242	0.004	0.454	0.496	0.513	0.000
1250	3000	50	0.001	1.477	0.054	0.000	0.058	0.015	1224.342	0.499	0.002	-75013.808	0.006	0.449	0.487	0.511	0.000
625	3000	50	-0.001	1.481	0.055	0.001	0.053	0.015	1223.860	0.501	0.001	-74771.204	0.009	0.451	0.483	0.518	0.000
312	3000	50	-0.001	1.477	0.055	0.001	0.053	0.015	1226.967	0.480	0.002	-74905.122	0.012	0.430	0.455	0.505	0.001
156	3000	50	0.001	1.473	0.054	-0.001	0.061	0.014	1229.901	0.464	0.002	-75123.100	0.017	0.414	0.429	0.499	0.001
78	3000	50	0.002	1.473	0.054	-0.001	0.061	0.015	1220.230	0.526	0.001	-74769.185	0.025	0.476	0.477	0.575	0.003
20	3000	50	-0.003	1.484	0.053	0.004	0.044	0.012	1232.012	0.464	0.002	-75249.604	0.049	0.414	0.367	0.562	0.011
# of replication	sample size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	50	0.000	1.478	0.067	0.000	0.123	0.018	1223.924	0.502	0.002	-49959.823	0.003	0.452	0.496	0.508	0.000
2500	2000	50	0.001	1.476	0.067	0.000	0.123	0.018	1223.882	0.501	0.002	-49988.694	0.004	0.451	0.492	0.510	0.000
1250	2000	50	0.002	1.485	0.067	-0.001	0.126	0.018	1223.874	0.500	0.002	-49928.566	0.006	0.450	0.488	0.513	0.000
625	2000	50	0.000	1.480	0.067	0.001	0.123	0.018	1222.493	0.507	0.002	-49963.840	0.009	0.457	0.490	0.524	0.000
312	2000	50	0.000	1.479	0.067	0.002	0.056	0.017	1222.136	0.506	0.002	-49932.391	0.012	0.456	0.481	0.530	0.001
156	2000	50	0.002	1.490	0.068	-0.002	0.118	0.017	1222.925	0.510	0.002	-49758.711	0.017	0.460	0.475	0.545	0.001
78	2000	50	0.001	1.477	0.066	-0.001	0.126	0.019	1213.061	0.558	0.002	-50098.217	0.025	0.508	0.509	0.608	0.003
20	2000	50	-0.015	1.467	0.067	0.013	0.030	0.021	1242.775	0.411	0.003	-50266.496	0.049	0.361	0.313	0.508	0.011
# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	50	-0.001	1.482	0.095	0.002	0.145	0.025	1225.421	0.491	0.003	-24952.975	0.003	0.441	0.484	0.497	0.000
2500	1000	50	0.000	1.492	0.096	0.000	0.074	0.026	1225.135	0.498	0.003	-24924.722	0.004	0.448	0.489	0.506	0.000
1250	1000	50	0.001	1.484	0.095	0.000	0.072	0.025	1222.830	0.505	0.003	-24938.207	0.006	0.455	0.492	0.517	0.000
625	1000	50	-0.001	1.485	0.095	0.001	0.071	0.026	1224.304	0.499	0.003	-24909.186	0.009	0.449	0.482	0.517	0.000
312	1000	50	0.002	1.485	0.095	-0.001	0.073	0.023	1227.370	0.477	0.003	-24949.250	0.012	0.427	0.452	0.502	0.001
156	1000	50	-0.002	1.489	0.095	0.004	0.064	0.024	1220.446	0.523	0.002	-24928.376	0.017	0.473	0.488	0.558	0.001
78	1000	50	-0.001	1.483	0.093	0.002	0.069	0.027	1237.680	0.426	0.003	-24993.260	0.025	0.376	0.376	0.475	0.003
20	1000	50	-0.013	1.484	0.098	0.013	0.054	0.034	1214.549	0.539	0.002	-24920.461	0.049	0.489	0.441	0.636	0.011
# of replication	Sample.size	# of item	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	50	0.002	1.489	0.135	0.000	0.087	0.036	1226.638	0.486	0.004	-12468.483	0.003	0.436	0.480	0.492	0.000
2500	500	50	0.003	1.487	0.135	-0.001	0.089	0.036	1223.561	0.483	0.004	-12437.121	0.004	0.433	0.474	0.491	0.000
1250	500	50	0.002	1.493	0.135	0.001	0.085	0.036	1224.014	0.489	0.004	-12424.144	0.006	0.439	0.476	0.501	0.000
625	500	50	0.004	1.480	0.133	-0.001	0.090	0.035	1217.708	0.489	0.004	-12376.198	0.009	0.439	0.471	0.506	0.000
312	500	50	0.001	1.489	0.135	0.002	0.091	0.040	1227.775	0.487	0.004	-12461.296	0.012	0.437	0.462	0.511	0.001
156	500	50	0.002	1.495	0.135	0.001	0.087	0.037	1217.750	0.491	0.004	-12340.218	0.017	0.441	0.456	0.526	0.001
78	500	50	-0.001	1.521	0.138	0.002	0.082	0.033	1221.873	0.516	0.003	-12380.931	0.025	0.466	0.466	0.565	0.003
20	500	50	-0.002	1.510	0.139	0.009	0.066	0.034	1231.039	0.445	0.004	-12343.249	0.049	0.395	0.348	0.543	0.011

7. 2. Supplementary File for 2PL Model

# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	3000	25	0.004	0.163	0.085	0.003	1.458	0.075	-0.001	0.110	0.014	275.439	0.496	0.002	-35127.778	0.003	0.446	0.490	0.502	0.000
2500	3000	25	0.004	0.162	0.085	0.003	1.457	0.075	-0.001	0.111	0.014	275.0509	0.500	0.002	-35152.720	0.004	0.450	0.491	0.508	0.000
1250	3000	25	0.004	0.162	0.085	0.002	1.465	0.076	0.000	0.114	0.015	275.7612	0.493	0.002	-35107.61051	0.006	0.443	0.481	0.506	0.000
625	3000	25	0.004	0.164	0.086	0.004	1.469	0.076	-0.001	0.114	0.014	273.6207	0.519	0.002	-35123.35379	0.009	0.469	0.502	0.536	0.000
312	3000	25	0.005	0.162	0.085	0.005	1.462	0.076	-0.001	0.112	0.014	275.6247	0.496	0.002	-35110.32823	0.012	0.446	0.472	0.521	0.001
156	3000	25	0.003	0.164	0.085	0.000	1.467	0.074	0.000	0.109	0.014	277.8721	0.476	0.003	-35033.99359	0.017	0.426	0.441	0.511	0.001
78	3000	25	0.003	0.162	0.086	-0.002	1.451	0.076	0.002	0.111	0.013	272.1231	0.528	0.002	-35076.88341	0.025	0.478	0.478	0.577	0.003
20	3000	25	0.002	0.167	0.086	-0.009	1.479	0.073	0.003	0.095	0.013	279.572	0.453	0.003	-35204.453	0.049	0.403	0.356	0.550	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	25	0.005	0.172	0.105	0.003	1.463	0.092	0.000	0.061	0.018	275.122	0.499	0.003	-23400.047	0.003	0.449	0.493	0.505	0.000
2500	2000	25	0.004	0.172	0.105	0.004	1.462	0.094	-0.001	0.062	0.018	275.495	0.492	0.003	-23421.325	0.004	0.442	0.483	0.501	0.000
1250	2000	25	0.005	0.172	0.105	0.005	1.460	0.094	-0.002	0.064	0.018	275.768	0.490	0.003	-23418.543	0.006	0.440	0.478	0.502	0.000
625	2000	25	0.004	0.172	0.104	0.003	1.443	0.091	-0.001	0.061	0.018	277.119	0.475	0.003	-23474.757	0.009	0.425	0.457	0.492	0.000
312	2000	25	0.002	0.170	0.105	0.004	1.467	0.094	-0.001	0.062	0.019	275.991	0.489	0.003	-23415.020	0.012	0.439	0.465	0.514	0.001
156	2000	25	0.006	0.175	0.108	0.004	1.484	0.097	-0.001	0.061	0.018	276.621	0.483	0.003	-23351.905	0.017	0.433	0.448	0.518	0.001
78	2000	25	0.002	0.171	0.102	0.002	1.451	0.089	-0.002	0.061	0.016	275.571	0.501	0.003	-23463.769	0.025	0.451	0.452	0.551	0.003
20	2000	25	0.002	0.161	0.102	0.007	1.482	0.087	-0.003	0.061	0.015	287.456	0.325	0.004	-23384.685	0.049	0.275	0.228	0.423	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	25	0.010	0.199	0.149	0.005	1.470	0.132	-0.001	0.146	0.025	275.917	0.489	0.004	-11688.028	0.003	0.439	0.483	0.495	0.000
2500	1000	25	0.010	0.199	0.149	0.004	1.468	0.132	0.000	0.148	0.025	276.2027	0.486	0.004	-11687.15335	0.004	0.436	0.477	0.494	0.000
1250	1000	25	0.012	0.198	0.149	0.004	1.463	0.132	0.000	0.145	0.025	274.7629	0.502	0.004	-11697.04832	0.006	0.452	0.490	0.515	0.000
625	1000	25	0.010	0.201	0.150	0.008	1.460	0.137	-0.002	0.145	0.026	273.7744	0.511	0.004	-11725.39484	0.009	0.461	0.494	0.529	0.000
312	1000	25	0.014	0.198	0.148	0.001	1.466	0.133	0.001	0.148	0.027	275.9167	0.490	0.004	-11702.90608	0.012	0.440	0.465	0.514	0.001
156	1000	25	0.009	0.201	0.147	0.010	1.467	0.133	-0.005	0.154	0.027	276.600	0.487	0.004	-11687.921	0.017	0.437	0.452	0.522	0.001
78	1000	25	0.016	0.198	0.148	-0.002	1.440	0.129	0.004	0.162	0.030	273.579	0.509	0.004	-11754.848	0.025	0.459	0.460	0.558	0.003
20	1000	25	0.007	0.207	0.150	-0.002	1.458	0.139	0.002	0.125	0.021	274.042	0.502	0.004	-11755.487	0.049	0.452	0.405	0.599	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	25	0.022	0.249	0.215	0.008	1.481	0.191	0.000	0.087	0.036	275.819	0.491	0.005	-5838.883	0.003	0.441	0.485	0.498	0.000
2500	500	25	0.022	0.248	0.215	0.009	1.480	0.191	-0.001	0.087	0.036	276.222	0.486	0.006	-5836.567	0.004	0.436	0.477	0.495	0.000
1250	500	25	0.022	0.249	0.216	0.010	1.484	0.191	-0.001	0.086	0.035	277.128	0.476	0.006	-5834.162	0.006	0.426	0.463	0.488	0.000
625	500	25	0.025	0.250	0.217	0.009	1.497	0.193	0.001	0.084	0.036	276.863	0.477	0.006	-5820.697	0.009	0.427	0.459	0.494	0.000
312	500	25	0.022	0.251	0.216	0.005	1.466	0.193	0.002	0.084	0.037	274.879	0.496	0.006	-5848.941	0.012	0.446	0.472	0.521	0.001
156	500	25	0.029	0.249	0.220	0.008	1.464	0.204	0.002	0.075	0.034	279.230	0.452	0.006	-5861.134	0.017	0.402	0.417	0.487	0.001
78	500	25	0.029	0.244	0.212	0.007	1.472	0.191	0.002	0.087	0.032	279.297	0.451	0.006	-5856.095	0.025	0.401	0.402	0.500	0.003
20	500	25	0.0177	0.256	0.214	-0.01	1.517	0.196	0.005	0.066	0.036	270.278	0.545	0.005	-5753.086	0.049	0.495	0.447	0.642	0.011

# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
4232	3000	50	0.003	0.163	0.082	0.003	1.482	0.075	-0.001	0.110	0.014	1176.523	0.49	0.002	-67642.223	0.003	0.440	0.483	0.497	0.000
2500	3000	50	0.003	0.163	0.081	0.004	1.485	0.074	-0.002	0.110	0.014	1176.306	0.491	0.002	-67603.514	0.004	0.441	0.482	0.500	0.000
1250	3000	50	0.003	0.162	0.082	0.004	1.485	0.074	-0.001	0.111	0.015	1177.651	0.485	0.002	-67678.284	0.006	0.435	0.473	0.497	0.000
625	3000	50	0.004	0.164	0.082	0.002	1.491	0.075	-0.001	0.112	0.014	1178.235	0.483	0.002	-67532.975	0.009	0.433	0.466	0.500	0.000
312	3000	50	0.003	0.163	0.082	0.002	1.476	0.076	-0.001	0.111	0.014	1172.113	0.515	0.001	-67731.533	0.012	0.465	0.490	0.540	0.001
156	3000	50	0.003	0.164	0.081	-0.001	1.482	0.075	0.002	0.109	0.015	1184.222	0.455	0.002	-67716.637	0.017	0.405	0.420	0.490	0.001
78	3000	50	0.000	0.166	0.083	-0.002	1.492	0.078	0.002	0.096	0.013	1172.677	0.513	0.002	-67576.411	0.025	0.463	0.464	0.562	0.003
20	3000	50	0.006	0.163	0.082	-0.015	1.479	0.085	0.008	0.144	0.020	1192.958	0.419	0.002	-67489.537	0.049	0.369	0.322	0.516	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	50	0.005	0.172	0.100	0.005	1.483	0.091	-0.002	0.062	0.018	1177.890	0.483	0.002	-45105.882	0.003	0.433	0.477	0.489	0.000
2500	2000	50	0.005	0.172	0.100	0.005	1.481	0.091	-0.002	0.062	0.018	1178.677	0.478	0.002	-45114.507	0.004	0.428	0.469	0.487	0.000
1250	2000	50	0.005	0.172	0.100	0.004	1.482	0.092	-0.002	0.061	0.018	1180.331	0.467	0.002	-45107.194	0.006	0.417	0.454	0.479	0.000
625	2000	50	0.003	0.171	0.100	0.007	1.484	0.091	-0.003	0.067	0.018	1179.662	0.476	0.002	-45083.287	0.009	0.426	0.459	0.494	0.000
312	2000	50	0.006	0.172	0.102	0.010	1.499	0.094	-0.004	0.070	0.018	1173.844	0.509	0.002	-44964.306	0.012	0.459	0.484	0.533	0.001
156	2000	50	0.008	0.172	0.100	0.003	1.479	0.090	0.000	0.057	0.018	1185.649	0.428	0.002	-45051.152	0.017	0.378	0.393	0.463	0.001
78	2000	50	0.000	0.171	0.100	-0.005	1.455	0.093	0.003	0.053	0.020	1173.769	0.492	0.002	-45380.808	0.025	0.442	0.443	0.541	0.003
20	2000	50	0.004	0.165	0.097	0.006	1.472	0.085	-0.004	0.061	0.016	1171.665	0.531	0.001	-45367.346	0.049	0.481	0.433	0.628	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	50	0.010	0.198	0.143	0.004	1.491	0.131	0.000	0.146	0.025	1178.332	0.481	0.003	-22521.259	0.003	0.431	0.475	0.487	0.000
2500	1000	50	0.009	0.198	0.144	0.005	1.491	0.130	-0.001	0.145	0.025	1180.073	0.470	0.003	-22523.647	0.004	0.420	0.461	0.479	0.000
1250	1000	50	0.010	0.197	0.143	0.007	1.488	0.131	-0.002	0.149	0.025	1178.503	0.480	0.003	-22511.257	0.006	0.430	0.468	0.492	0.000
625	1000	50	0.008	0.198	0.143	0.001	1.495	0.130	0.002	0.143	0.025	1179.117	0.474	0.003	-22520.229	0.009	0.424	0.457	0.491	0.000
312	1000	50	0.010	0.199	0.144	0.005	1.495	0.133	-0.001	0.145	0.026	1171.933	0.522	0.003	-22480.886	0.012	0.472	0.497	0.547	0.001
156	1000	50	0.007	0.196	0.141	0.011	1.486	0.127	-0.003	0.146	0.022	1173.195	0.512	0.002	-22520.435	0.017	0.462	0.477	0.547	0.001
78	1000	50	0.003	0.196	0.141	-0.003	1.499	0.131	0.002	0.143	0.027	1173.711	0.510	0.003	-22478.467	0.025	0.460	0.461	0.559	0.003
20	1000	50	0.011	0.196	0.146	-0.018	1.411	0.126	0.012	0.128	0.027	1194.590	0.379	0.004	-22841.479	0.049	0.329	0.282	0.476	0.011
# of replication	Sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq.5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	50	0.020	0.244	0.207	0.009	1.509	0.190	-0.001	0.087	0.036	1180.094	0.471	0.004	-11231.536	0.003	0.421	0.465	0.478	0.000
2500	500	50	0.021	0.244	0.207	0.011	1.507	0.192	-0.002	0.089	0.037	1179.945	0.471	0.004	-11228.371	0.004	0.421	0.463	0.480	0.000
1250	500	50	0.018	0.243	0.206	0.011	1.507	0.192	-0.002	0.088	0.037	1179.010	0.478	0.004	-11242.831	0.006	0.428	0.466	0.490	0.000
625	500	50	0.020	0.244	0.207	0.003	1.505	0.190	0.003	0.076	0.035	1179.340	0.477	0.004	-11244.709	0.009	0.427	0.460	0.495	0.000
312	500	50	0.020	0.244	0.206	0.004	1.489	0.187	0.002	0.082	0.036	1172.149	0.516	0.004	-11243.590	0.012	0.466	0.492	0.541	0.001
156	500	50	0.024	0.245	0.208	0.004	1.510	0.191	0.001	0.080	0.037	1173.253	0.513	0.003	-11216.526	0.017	0.463	0.478	0.548	0.001
78	500	50	0.021	0.242	0.210	0.009	1.516	0.189	-0.001	0.086	0.033	1179.763	0.476	0.004	-11173.228	0.025	0.426	0.427	0.525	0.003
20	500	50	0.003	0.248	0.215	0.037	1.529	0.199	-0.017	0.122	0.033	1186.695	0.426	0.005	-11300.049	0.049	0.376	0.329	0.524	0.011

7. 3. Supplementary File for 3PL Model

# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	3000	25	0.021	0.232	0.191	-0.004	1.463	0.203	-0.002	0.083	0.077	0.000	0.056	0.014	249.045	0.512	0.002	-38094.926	0.003	0.462	0.506	0.518	0.000
2500	3000	25	0.021	0.231	0.191	-0.004	1.459	0.202	-0.002	0.082	0.077	0.000	0.055	0.015	248.907	0.516	0.002	-38086.024	0.004	0.466	0.507	0.524	0.000
1250	3000	25	0.021	0.233	0.192	-0.004	1.470	0.204	-0.002	0.082	0.076	-0.001	0.058	0.015	250.148	0.495	0.002	-38103.522	0.006	0.445	0.482	0.507	0.000
625	3000	25	0.021	0.231	0.192	-0.005	1.458	0.204	-0.001	0.082	0.076	0.000	0.055	0.015	249.078	0.510	0.002	-38167.074	0.009	0.460	0.492	0.527	0.000
312	3000	25	0.022	0.234	0.191	-0.004	1.471	0.206	-0.001	0.083	0.077	0.000	0.057	0.015	249.633	0.505	0.002	-38195.932	0.012	0.455	0.480	0.530	0.001
156	3000	25	0.021	0.233	0.190	-0.002	1.440	0.203	-0.001	0.081	0.075	-0.001	0.063	0.016	249.855	0.494	0.002	-38276.803	0.017	0.444	0.459	0.529	0.001
78	3000	25	0.021	0.225	0.184	-0.014	1.413	0.200	0.001	0.081	0.076	0.003	0.047	0.014	253.372	0.455	0.003	-37804.962	0.025	0.405	0.406	0.504	0.003
20	3000	25	0.008	0.229	0.194	-0.002	1.470	0.228	-0.007	0.082	0.076	0.004	0.035	0.012	255.648	0.421	0.003	-38432.708	0.049	0.371	0.324	0.519	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	25	0.034	0.271	0.240	-0.009	1.475	0.251	-0.001	0.093	0.089	0.000	0.123	0.018	248.774	0.515	0.003	-25390.889	0.003	0.465	0.509	0.521	0.000
2500	2000	25	0.034	0.270	0.239	-0.008	1.477	0.250	-0.002	0.093	0.089	0.000	0.124	0.018	248.317	0.521	0.002	-25313.220	0.004	0.471	0.512	0.530	0.000
1251	2000	25	0.035	0.270	0.238	-0.011	1.470	0.250	-0.001	0.092	0.088	-0.001	0.122	0.018	248.501	0.521	0.002	-25373.037	0.006	0.471	0.509	0.534	0.000
625	2000	25	0.033	0.269	0.238	-0.009	1.473	0.249	-0.001	0.093	0.089	-0.001	0.124	0.018	248.238	0.521	0.002	-25411.342	0.009	0.471	0.503	0.538	0.000
312	2000	25	0.035	0.273	0.244	-0.010	1.487	0.255	-0.002	0.096	0.092	0.000	0.122	0.017	250.301	0.493	0.003	-25113.030	0.012	0.443	0.469	0.518	0.001
156	2000	25	0.042	0.271	0.237	-0.018	1.491	0.249	0.000	0.093	0.088	0.000	0.138	0.019	248.732	0.522	0.003	-25393.005	0.017	0.472	0.487	0.557	0.001
78	2000	25	0.040	0.281	0.248	-0.019	1.500	0.249	0.001	0.097	0.092	-0.001	0.124	0.017	246.969	0.549	0.002	-25333.664	0.025	0.499	0.499	0.598	0.003
20	2000	25	0.039	0.276	0.244	-0.014	1.419	0.267	0.004	0.093	0.087	0.001	0.092	0.010	243.860	0.593	0.001	-25785.769	0.049	0.543	0.496	0.691	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	25	0.074	0.374	0.363	-0.024	1.506	0.362	-0.001	0.113	0.111	0.000	0.146	0.025	249.099	0.512	0.004	-12675.636	0.003	0.462	0.505	0.518	0.000
2500	1000	25	0.077	0.383	0.371	-0.026	1.516	0.370	0.001	0.113	0.111	0.000	0.147	0.025	248.589	0.520	0.004	-12675.481	0.004	0.470	0.511	0.529	0.000
1250	1000	25	0.076	0.382	0.371	-0.025	1.509	0.368	0.000	0.114	0.112	0.001	0.148	0.025	248.443	0.522	0.004	-12674.330	0.006	0.472	0.510	0.535	0.000
625	1000	25	0.073	0.372	0.359	-0.024	1.501	0.362	-0.001	0.113	0.111	0.002	0.146	0.026	248.465	0.519	0.004	-12671.896	0.009	0.469	0.501	0.536	0.000
312	1000	25	0.072	0.381	0.369	-0.023	1.533	0.365	-0.001	0.116	0.114	0.001	0.142	0.024	247.811	0.528	0.003	-12607.724	0.012	0.478	0.503	0.552	0.001
156	1000	25	0.073	0.388	0.375	-0.014	1.523	0.374	0.001	0.112	0.110	-0.001	0.147	0.025	248.867	0.507	0.004	-12735.953	0.017	0.457	0.472	0.541	0.001
78	1000	25	0.071	0.374	0.361	-0.029	1.469	0.347	0.000	0.112	0.110	0.000	0.153	0.028	250.695	0.498	0.004	-12605.100	0.025	0.448	0.449	0.548	0.003
20	1000	25	0.073	0.370	0.352	-0.031	1.469	0.390	-0.001	0.110	0.110	-0.006	0.180	0.030	243.709	0.575	0.002	-12792.127	0.049	0.525	0.477	0.672	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	25	0.179	0.631	0.649	-0.050	1.621	0.588	0.002	0.141	0.141	-0.001	0.088	0.035	247.992	0.526	0.005	-6309.552	0.003	0.476	0.520	0.532	0.000
2500	500	25	0.182	0.635	0.652	-0.052	1.622	0.589	0.002	0.140	0.140	0.000	0.085	0.035	248.248	0.520	0.005	-6309.280	0.004	0.470	0.511	0.528	0.000
1250	500	25	0.175	0.630	0.645	-0.053	1.634	0.601	0.002	0.141	0.141	0.000	0.087	0.037	246.302	0.541	0.005	-6300.432	0.006	0.491	0.528	0.553	0.000
625	500	25	0.173	0.614	0.633	-0.050	1.597	0.578	0.002	0.140	0.140	-0.001	0.086	0.035	249.147	0.508	0.005	-6318.462	0.009	0.458	0.490	0.525	0.000
312	500	25	0.193	0.626	0.647	-0.065	1.617	0.596	0.002	0.140	0.141	-0.002	0.088	0.036	249.015	0.513	0.005	-6336.257	0.012	0.463	0.488	0.537	0.001
156	500	25	0.195	0.646	0.670	-0.032	1.594	0.580	0.003	0.145	0.145	-0.004	0.096	0.038	246.692	0.542	0.005	-6249.925	0.017	0.492	0.507	0.577	0.001
78	500	25	0.199	0.694	0.714	-0.055	1.663	0.602	0.003	0.141	0.141	0.004	0.080	0.040	241.654	0.592	0.004	-6294.088	0.025	0.542	0.543	0.642	0.003
20	500	25	0.201	0.703	0.735	-0.030	1.731	0.628	-0.001	0.144	0.141	0.012	0.060	0.031	250.054	0.501	0.006	-6342.525	0.049	0.451	0.403	0.598	0.011

# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	3000	50	0.016	0.217	0.168	-0.003	1.489	0.185	-0.002	0.078	0.071	-0.001	0.056	0.014	1122.474	0.517	0.001	-74079.570	0.003	0.467	0.510	0.523	0.000
2500	3000	50	0.016	0.218	0.168	-0.002	1.487	0.185	-0.002	0.078	0.071	0.000	0.054	0.015	1122.125	0.519	0.001	-74090.428	0.004	0.469	0.510	0.527	0.000
1250	3000	50	0.018	0.217	0.168	-0.004	1.489	0.185	-0.001	0.077	0.070	0.000	0.056	0.015	1122.678	0.517	0.001	-74100.260	0.006	0.467	0.504	0.529	0.000
625	3000	50	0.021	0.217	0.169	-0.004	1.488	0.183	-0.001	0.078	0.071	-0.001	0.058	0.015	1120.749	0.527	0.001	-73851.830	0.009	0.477	0.510	0.544	0.000
312	3000	50	0.015	0.219	0.169	-0.001	1.490	0.184	-0.002	0.079	0.071	0.000	0.053	0.014	1126.778	0.491	0.002	-74093.610	0.012	0.441	0.466	0.515	0.001
156	3000	50	0.014	0.218	0.169	-0.004	1.495	0.185	-0.002	0.081	0.074	-0.001	0.059	0.014	1120.568	0.530	0.001	-73702.999	0.017	0.480	0.495	0.565	0.001
78	3000	50	0.010	0.216	0.166	-0.004	1.466	0.184	-0.002	0.079	0.071	0.001	0.053	0.014	1125.253	0.498	0.001	-74269.253	0.025	0.448	0.448	0.547	0.003
20	3000	50	0.019	0.211	0.163	-0.005	1.450	0.172	-0.004	0.078	0.072	0.003	0.047	0.015	1134.891	0.445	0.002	-73364.730	0.049	0.395	0.348	0.543	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	2000	50	0.029	0.250	0.211	-0.010	1.496	0.230	-0.001	0.088	0.082	0.000	0.124	0.018	1122.660	0.514	0.002	-49391.158	0.003	0.464	0.508	0.520	0.000
2500	2000	50	0.028	0.250	0.211	-0.009	1.498	0.230	-0.001	0.088	0.082	0.000	0.122	0.018	1123.245	0.510	0.002	-49394.856	0.004	0.460	0.501	0.519	0.000
1250	2000	50	0.029	0.250	0.212	-0.011	1.490	0.229	0.000	0.088	0.082	0.000	0.124	0.018	1122.263	0.518	0.002	-49411.865	0.006	0.468	0.505	0.530	0.000
625	2000	50	0.027	0.249	0.210	-0.008	1.490	0.230	0.000	0.088	0.082	-0.001	0.126	0.018	1124.352	0.500	0.002	-49411.525	0.009	0.450	0.483	0.518	0.000
312	2000	50	0.029	0.250	0.212	-0.011	1.492	0.230	0.000	0.089	0.083	-0.002	0.123	0.017	1120.916	0.533	0.002	-49393.274	0.012	0.483	0.508	0.558	0.001
156	2000	50	0.026	0.248	0.209	-0.012	1.516	0.230	-0.001	0.088	0.082	0.002	0.114	0.017	1123.579	0.511	0.002	-49213.810	0.017	0.461	0.476	0.546	0.001
78	2000	50	0.031	0.251	0.216	-0.009	1.498	0.229	-0.002	0.088	0.083	0.001	0.126	0.015	1121.690	0.499	0.002	-49426.222	0.025	0.449	0.450	0.549	0.003
20	2000	50	0.039	0.253	0.215	-0.025	1.540	0.217	0.003	0.090	0.083	0.000	0.121	0.016	1115.446	0.564	0.002	-49194.058	0.049	0.514	0.467	0.662	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	1000	50	0.062	0.335	0.315	-0.022	1.517	0.325	0.001	0.107	0.104	0.000	0.073	0.025	1121.134	0.523	0.002	-24619.555	0.003	0.473	0.517	0.530	0.000
2500	1000	50	0.063	0.336	0.316	-0.019	1.523	0.328	0.001	0.107	0.103	-0.002	0.077	0.025	1119.738	0.530	0.002	-24626.171	0.004	0.480	0.521	0.539	0.000
1251	1000	50	0.061	0.338	0.317	-0.019	1.521	0.328	0.001	0.107	0.103	-0.001	0.147	0.025	1122.340	0.515	0.003	-24637.558	0.006	0.465	0.502	0.527	0.000
625	1000	50	0.064	0.339	0.318	-0.025	1.524	0.329	0.001	0.107	0.103	0.000	0.148	0.025	1119.339	0.531	0.002	-24603.283	0.009	0.481	0.513	0.548	0.000
312	1000	50	0.068	0.338	0.317	-0.029	1.535	0.334	0.001	0.107	0.103	0.000	0.149	0.026	1117.253	0.543	0.002	-24645.979	0.012	0.493	0.518	0.567	0.001
156	1000	50	0.066	0.330	0.309	-0.017	1.520	0.321	0.001	0.107	0.103	-0.002	0.147	0.026	1121.656	0.525	0.002	-24488.165	0.017	0.475	0.490	0.560	0.001
78	1000	50	0.059	0.329	0.310	-0.025	1.489	0.325	0.002	0.105	0.100	0.004	0.136	0.026	1121.085	0.518	0.002	-24768.173	0.025	0.468	0.469	0.568	0.003
20	1000	50	0.064	0.339	0.319	-0.027	1.506	0.340	-0.004	0.104	0.100	0.006	0.146	0.029	1119.694	0.531	0.002	-24831.287	0.049	0.481	0.434	0.629	0.011
# of replication	sample.size	# of item	me.a	se.a	rmse.a	me.b	se.b	rmse.b	me.c	se.c	rmse.c	me.theta	se.theta	rmse.theta	M2	M2.p	RMSEA.2	Log-Likelihood	Eq. 5	empirical-nominal alpha	-2	2	accuracy of error rate estimate
5000	500	50	0.146	0.538	0.544	-0.047	1.602	0.519	0.004	0.133	0.131	-0.001	0.088	0.037	1119.046	0.531	0.003	-12272.765	0.003	0.481	0.524	0.537	0.000
2500	500	50	0.143	0.541	0.546	-0.047	1.598	0.521	0.004	0.133	0.131	-0.001	0.086	0.036	1119.896	0.526	0.003	-12278.759	0.004	0.476	0.517	0.535	0.000
1250	500	50	0.143	0.540	0.546	-0.045	1.612	0.528	0.003	0.132	0.130	-0.001	0.086	0.035	1122.462	0.519	0.004	-12271.311	0.006	0.469	0.507	0.532	0.000
625	500	50	0.141	0.538	0.543	-0.051	1.591	0.522	0.003	0.133	0.131	0.001	0.078	0.033	1121.053	0.521	0.003	-12277.678	0.009	0.471	0.503	0.538	0.000
312	500	50	0.139	0.523	0.525	-0.050	1.594	0.512	0.004	0.133	0.131	0.000	0.081	0.036	1118.110	0.539	0.003	-12309.872	0.012	0.489	0.515	0.564	0.001
156	500	50	0.146	0.567	0.571	-0.050	1.594	0.518	0.004	0.133	0.131	0.005	0.072	0.038	1125.730	0.501	0.004	-12288.918	0.017	0.451	0.466	0.536	0.001
78	500	50	0.139	0.524	0.528	-0.052	1.616	0.539	0.003	0.131	0.130	-0.003	0.093	0.039	1117.142	0.552	0.003	-12288.143	0.025	0.502	0.502	0.601	0.003
20	500	50	0.168	0.583	0.592	-0.046	1.668	0.553	0.001	0.134	0.134	0.007	0.071	0.038	1125.758	0.467	0.004	-12256.452	0.049	0.417	0.369	0.564	0.011