International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I1P112

Volume 72 Issue 1, 117-129, January 2024

Original Article

The Joint Model of Two-Parameter Logistic and Response Time Model for Computer-Based Tests

Ninik Zuroidah[1,2], Kumaidi[3], Samsul Hadi[4], Kusaeri[5], Syahrul Ramadhan[6]

[1,4]*Yogyakarta State University, Sleman, Indonesia.*
[2]*Institute Agama Islam Negeri Kediri, Kediri, Indonesia.*
[3]*University Muhammadiyah Surakarta, Sukoharjo, Indonesia.*
[5]*UIN Sunan Ampel, Surabaya, Indonesia.*
[6]*National Research and Innovation Agency, Jakarta, Indonesia.*

[6]*Corresponding Author: syah030@brin.go.id*

Received: 15 August 2023 Revised: 18 November 2023 Accepted: 18 December 2023 Published: 07 January 2024

Abstract *- The purpose of this research is to produce a test measurement model that can explain the realistic conditions of the computer-based test by considering the response time. This research is a measurement model development consisting of two main analyses, such as 1) developing a joint model of two-parameter logistic and response time model and 2) examining the accuracy of the parameter model by using standard deviation and testing model fit by using DIC statistics. The model is implemented on empirical data from the National Selection test results of 1559 New Learners of Madrasah Aliyah Negeri (MAN) Insan Cendekia Ministry of Religion of the Republic of Indonesia in 2019 in the field of Mathematics consisting of 15 items with a time limit of 45 minutes. Estimation of parameter models used the Bayesian MCMC method with R2WinBUGS Software. The results of this research show that the joint model of Two Parameter Logistic and response time model produces a hierarchical structure model with three (3) groups of parameters, i.e., person parameters, item parameters, and rho parameters. Measurement models involving response time can improve the accuracy of model parameters compared with Two Parameter Logistic models that are separated from response time. The empirical data in this research fit better with the joint model of Two Parameter Logistic with response time characterized by a smaller DIC value than the Two Parameter Logistic model separated by response time. Therefore, it proves that the joint model of the Two Parameter Logistic with response time is more suitable to be implemented for computer-based and time-constrained tests.*

Keywords *- Response accuracy, Response time, Two parameter logistic, Bayesian MCMC, CBT.*

1. Introduction

Computer-based tests can provide two important pieces of information in measuring the ability of test takers, i.e., information about Response Accuracy (RA) and Response Time (RT). The response accuracy of the test taker is typically represented as dichotomous data, indicating correct or incorrect answers to test items. On the other hand, response time is measured as continuous data [1, 2]. Response time shows how long the test taker could complete the test. In other words, it shows the range time between when the item is presented and when the item is answered by the test taker. Moreover, it is challenging to capture response time per item in paper and pencil tests. Yet, it is available in computer-based tests that could be an indicator to assess the ability of the test taker. The information on response time gives a description about the test taker's cognitive performance in processing the information to solve the problem in answering the test item. This case could help in understanding the real test condition, giving detailed feedback, and making a valid conclusion for intervention and improvement [3]. The duration of response time holds significant value for test administrators, particularly in assessments intended for selection purposes [1]. If there are two different test takers, the number of items answered correctly is the same, the test score is the same, but the time to answer the questions is different. Organizers can use the additional information of test takers' response time in making test conclusions, i.e., test takers pass quickly or pass slowly and fail the test quickly or fail the test slowly [4]. Response accuracy functions as an indicator assessment of the test taker's ability; meanwhile, response time has a role as a speed marker in answering the test [5]. In Item Response

Theory, so far, the estimation of the test taker's cognitive is based on the response accuracy without entering the information of response time [6, 7]. This case is due to the difficulty in obtaining the information on response time for each item test. Though IRT is considered suitable for assessing the ability in unlimited time tests [8], in fact, each test has its limited time to be answered [9]. The response pattern (response accuracy), response time, and self-confidence level are important factors in evaluating the test taker's cognitive [5, 10].

In a test having limited time, the test taker's behavior is reflected in a choice between a speed test and a power test. The faster it takes to answer the test questions, the lower the accuracy level; meanwhile, the longer it takes to answer, the higher the level of accuracy. This case shows that the response time affects the response accuracy. The speed that can be measured from the time record becomes the factor influencing the accuracy of the test response that has limited time. In this context, though the accuracy becomes an indicator of the test taker's ability, but the speed in answering the test also has an important role in the test taker's test result.

The study about the correlation between response accuracy and response speed has been researched for a long time but is often overlooked [11]. The Speed Ability Trade-off (SAT) theory suggests that test takers can manage their time in answering questions, and it affects the accuracy of their responses [2]. When prioritizing accuracy, the test taker has a high ability to choose a slower pace in answering compared to those aiming for speed. From this case, it highlights a negative association or inverse relationship between the speed and accuracy of responses from test takers. This dependency relationship between ability parameters and speed varies and is complex based on the level of observation.

At the individual level of observation, there is a negative correlation between speed and ability. But if the observation level changes in the population level, then it is plausible that test takers who work faster are candidates for higher ability, and test takers who work slower are candidates for lower ability [1, 12].

This suggests that at the population level, there is a positive correlation between ability and speed of answering test items. The correlation that occurs between ability and the speed of answering test items by test takers, both at the individual level and the population level, can be a very potential source of variation, minimizing the error in the measurement model to reduce measurement bias and improve the accuracy of parameter estimation results and increase the validity of measurement results.

The explanation shows the superiority of the test taker ability measurement model involving response time information, i.e., the measurement results are more accurate and able to explore the correlation between ability and speed parameters. But until now, the measurement of the ability of computer-based test takers has only used response accuracy information without involving response time information, so it is considered unable to reveal the real conditions of the test. On the other hand, computer-based tests make it very easy to obtain response time records for each test item for each test taker. Response time information can describe the behavior of test takers. Up until now, there are few researchers developing measurement models by jointly modeling response accuracy data and response time data to estimate test-taker and test item parameters and to explore the relationship between model parameters.

One of the studies was conducted by van der Linden, integrated response time information with response accuracy in a contemporary Item Response Theory (IRT) model. This study proposed a new model integrating accurate responses and response time as cognitive indicators in computer-based tests. However, the study conducted here has a novelty since the purpose of the study is to produce a test measurement model that can explain the realistic conditions of the test by considering the response time. The researcher is interested in studying and developing a test taker ability measurement model that involves response time information. The model is expected to be able to capture the real conditions of computer-based tests with limited answer time. Until now, no ability measurement model has been found with an ideal and standard response time for all conditions and is able to improve the validity of the test taker's ability measurement results and test item parameters.

Furthermore, this study will develop a framework model for measuring test taker ability by modeling response accuracy and response time. Modeling is done in stages or levels. At the first level, response accuracy is modeled by using the Two Parameter Logistic model to produce ability and speed parameters (person parameters). While the response time is modeled by using the Lognormal distribution model to produce test item difficulty and time intensity parameters (item parameters). At the second level, person parameters are jointly modeled together by using the Bivariate Normal Distribution Model to explain the correlation that occurs at the individual level between ability and speed parameters [13, 14]. Item parameters are jointly modeled using the Bivariate Normal distribution model to explain the correlation that occurs at the test item level, describing the relationship between the test item difficulty parameter and time intensity.

To find out the effect of response time on the measurement model in this study, the results of the estimation of person parameters and item parameters on the ability measurement model involving response time (joint model) will be compared to the accuracy of the ability measurement model that does not consider response time (separate model). Measurement models involving response time are organized hierarchically. The standard deviation value of each parameter of the joint model will be compared with the parameters in the separate model. The model parameters with higher accuracy are those with smaller standard deviation values.

Model parameter estimation in this study uses the Bayesian MCMC Gibbs Sampler approach. The Bayesian method is considered good if it is used to estimate measurement models with many complex parameters [15, 16]. The selection of the Bayesian method because has advantages, such as (1) it can estimate complex model parameters; (2) it is free from assumptions in the IRT model; (3) it can be applied to small samples [15]. The estimation of model parameters in this study used WinBUGS and R2WinBUGS software.

The results of the model development in this study will be applied to empirical data, the results of the CBT test of the national selection of new MAN Insan cendekia students of the Ministry of Religious Affairs of the Republic of Indonesia in 2019. To determine the size of the good model criteria, researchers use DIC (Deviance Information Criterion) statistics.

2. Method

2.1. Research Design

This research is development research. The purpose of this research is to develop a test-taker ability measurement model by including response time information in the joint models. The model framework is developed with a multilevel or hierarchical structure. This model can estimate all model parameters together while exploring the relationship patterns that occur between model parameters. The relationship between the model parameters becomes a source of variation in the measurement model, which has very potential to minimize the error or measurement error so that the hierarchical model can increase the accuracy of the model parameter estimation results and increase the validity of the test results.

2.2. Model Development Stage

The stages of model development are as follows: First, Researchers reviewed relevant previous research results, identified the need for model parameters, and studied the advantages and disadvantages of previous research results to get strong reasons for developing a new test-taker ability measurement model. Second, researchers chose a joint model approach with a hierarchical structure, and then they designed a new model framework by formulating mathematical model equations based on theoretical studies.

Third, Researchers formulated a mathematical model structure for response time modeling at the first level and modeling at the second level. Modeling at the first level, researchers model response accuracy data by using the Two Parameter Logistic model, which is useful for providing information about the accuracy of answers and the ability of test takers, as well as item characteristics. Researchers modeled response time data by using the Lognormal distribution model to produce information on the speed of answering and item characteristics. In this first stage of modeling, two groups of parameters are generated, such as person parameters and item parameters. Modeling at the second level, researchers formulated mathematical models to model person parameters and item parameters, resulting in 2 Bivariate Normal distribution model equations. Fourth, Researchers applied the model to empirical data.

Fifth, Researchers evaluated the model by testing the model's Goodness of Fit and looking at the accuracy of model parameters between measurement models involving response time (joint models of Two Parameter Logistic model and Lognormal model with a hierarchical structure) and measurement models without involving response time (Two Parameter Logistic modeled separately from Lognormal).

2.3. Research Subject

The empirical data used in this research is data on answer responses and response times of CBT test participants in the National Selection of New Learners (SNPDB) MAN UNGGULAN (Insan Cendekia) Ministry of Religion of the Republic of Indonesia in 2019 with a total of 1559 participants who worked on type A mathematics items from four types of mathematics questions, such as Type A, B, C, D. Each type of mathematics question consists of 15 items with a time limit of 45 minutes.

2.4. Data Analysis Techniques

Response accuracy data in the form of test takers' answers, i.e., correct answers scored 1, wrong answers scored 0. Meanwhile, response time data is a record of the length of time to answer test takers for each item. The length of time to answer in this study is measured starting from the time the test taker starts clicking on the item to be read answered until the participant clicks on the next item. This empirical data is used to estimate model parameters using the Bayesian Markov Chain Monte Carlo Gibbs Sampling method with the procedure: Determining the likelihood distribution according to the empirical data; Determining the prior distribution for each parameter in the model; Multiplying the likelihood distribution with the prior distribution to produce the posterior distribution; Generating parameter values in the posterior distribution using the Gibbs Sampling algorithm. To simplify the parameter estimation process in complex models, researchers use WinBUGS and R2WinBUGS software.

In this study, the accuracy of the ability measurement model parameters involving response time, namely the hierarchical joint model, will be compared with the ability measurement model parameters without involving response time, namely the separated model. The accuracy of the model parameters can be checked from the standard deviation value of each parameter. Model parameters that have a smaller standard deviation value. Then, the model parameters are more accurate. This is done to see the effect of response time on the parameters of the test-taker ability measurement model. The standard deviation equation is:

$$\sigma_\theta = \sqrt{\frac{\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2}{n}} \quad (1)$$

With is the standard deviation of the parameter and is the average value with n observations. Meanwhile, to choose the best model between the measurement model involving response time (hierarchical joint model) and the ability measurement model without involving response time (separate model), the researcher uses the DIC (Deviance Information Criterion) statistic with the equation as follows;

$$DIC = goodnessoffit + complexity \quad (2)$$

$$PD = E_{\theta|y}(D) - D\left(E_{\theta|y}(\theta)\right) = \bar{D} - D(\bar{\theta})$$

$$DIC = \bar{D} - PD \quad (3)$$

Equation (3) explains that the DIC statistic is the difference between the posterior mean deviation and the deviation in the posterior mean. PD is the estimated number of effective parameters [12, 17].

3. Results and Discussion

3.1. Joint Model of Two Parameter Logistics with Random Variable Response Time to Measure Test Takers' Ability

In this study, response time data is assumed to be a random variable, because each test taker has a different response time record for each item, as well as each test taker also has a different response time record for each item.

Response accuracy and response time are modeled together with a hierarchical structure in the test-taker ability measurement model. The following is a diagram of the modeling structure in this study, namely:

3.1.1. First Level Modeling

Researchers chose to use the Item Response Theory (IRT) 2 Parameter Logistic (PL) model to model response accuracy data by conducting a model fit test on empirical data using the chi-squared Test Statistic. Model fit test to determine the suitability (suitability) of empirical data test items with IRT 1 PL or, 2 PL, or 3 PL models. The decision on model fit is made by comparing the calculated chi-squared value with the chi-squared table value or by comparing the sig. (significance) of the test results with the researcher-determined test real level (α). If the sig. < α. Then, the test item is said not to fit one of the IRT models. and vice versa if sig. > α, then the test item is said to fit one of the IRT models [18]. The null hypothesis tested reads the data on the test items fits one of the IRT models, and for the alternative hypothesis reads the opposite. The calculation in this model fit test uses the help of R Studio software with the following results.

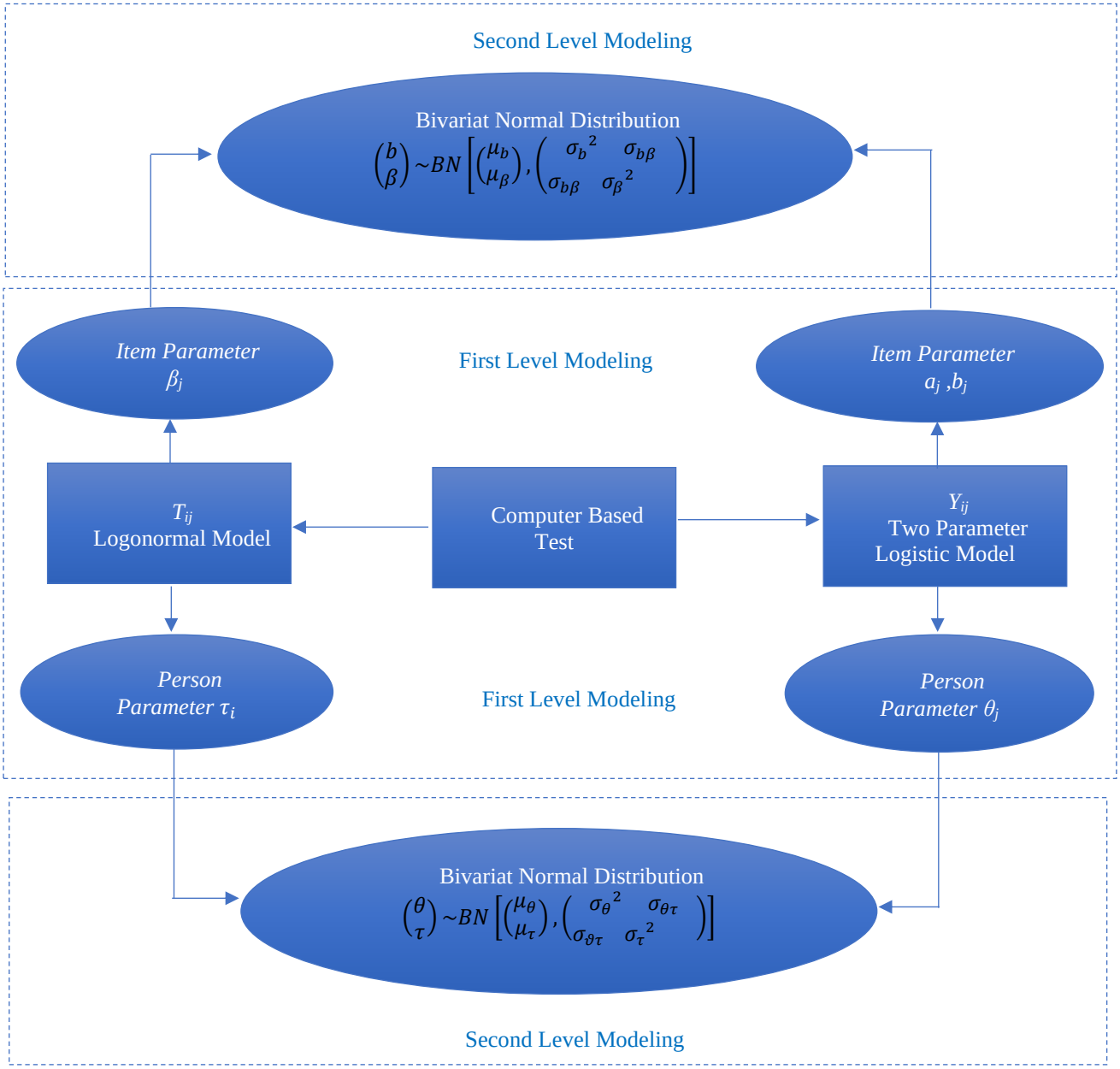


Fig. 1 Structure of the joint model of response accuracy and response time in hierarchy

Table 1 illustrates the comparison of the model fit test results using the chi-squared test statistic. For the empirical data fit test with the IRT 1 PL model, it is known that 15 test items have a sig. value greater than the 5% alpha value. This means that none of the 15 test items fit the IRT 1 PL model. For the empirical data fit test with the IRT 2 PL model, it is known that of the 15 test items, there are 9 test items that have a sig. value greater than the 5% alpha value and there are 6 test items whose sig. value is less than the 5% alpha value namely test item 5, item 6, item 7, item 8, item 12, and item 15. For the empirical data fit test with the IRT 3 PL model, it is known that of the 15 test items, there are 10 test items that have a sig. value greater than the 5% alpha value and there are 5 test items whose sig. value is less than the 5% alpha value, namely item 5, item 7, item 8, item 12 and item 15. This means that the model fit test results explain that the empirical data is more suitable for the IRT 2 PL model. The test conclusions are presented in Table 2 below.

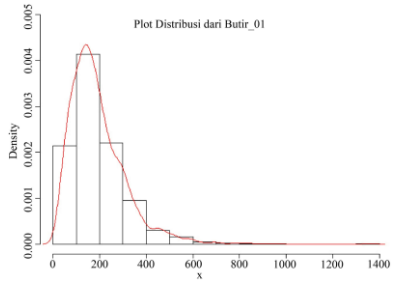
Table 2 shows that the empirical data distribution of test items is better suited to the IRT 2 PL model, so at the first level, the researcher uses the IRT 2 PL model to model the response accuracy data of the national selection results for the admission of new students of the flagship State Aliyah Madrasah of the Ministry of Religion of the Republic of Indonesia in 2019 with the number of test participants of 1559. The IRT 2 PL model is assumed to follow a two-parameter cumulative normal distribution model [19]. The estimation of the IRT 2 PL model using the Bayesian method resulted in the estimation of person parameters, namely the ability of test takers (θ_i) and item parameters (item difficulty (b_j) and item distinctiveness (a_j). Empirical data response time in this study has a value of more than zero. The following is a plot of the distribution of response time data on the results of the 2019 Kemenag RI superior MAN entrance exam using the help of R software. as follows:

Table 1. Model fit test results IRT 1 PL, 2 PL. and 3 PL with the response data of each item of test question package a in mathematics.

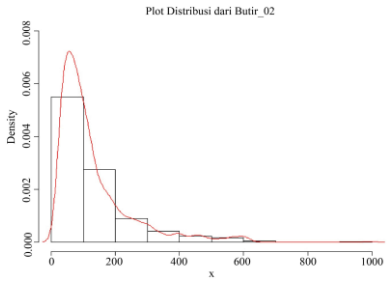
No, Item	1 PL		2 PL		3 PL	
	χ^2	Pr (>χ^2)	χ^2	χ^2	Pr (>χ^2)	χ^2
Item 01	104,2712	0,0001	590,0934	0,0001	584,3705	0,0001
Item 02	14,0918	0,015	14,8318	0,0001	148,5692	0,0001
Item 03	110,7858	0,0001	57,8916	0,0001	57,8253	0,0001
Item 04	55,7559	0,0001	46,3718	0,0001	42,9811	0,0001
Item 05	54,1977	0,0001	11,1024	0,1342	11,2222	0,1292
Item 06	65,4957	0,0001	15,0623	0,0579	15,9126	0,0259
Item 07	102,6633	0,0001	12,3439	0,0818	13,9917	0,0813
Item 08	60,6064	0,0001	75,843	0,3707	9,1230	0,2439
Item 09	81,6120	0,0001	70,4891	0,0001	74,7703	0,0001
Item 10	76,1878	0,0001	27,5375	0,0003	29,4229	0,0001
Item 11	63,6077	0,0001	30,0572	0,0001	29,5062	0,0001
Item 12	49,6258	0,0001	13,4247	0,0629	7,7643	0,3538
Item 13	93,8423	0,0001	23,2602	0,0001	23,6802	0,0001
Item 14	26,0130	0,0001	28,7585	0,0001	28,8672	0,0001
Item 15	65,1687	0,0001	14,1321	0,0689	13,9895	0,0814

Table 2. Conclusion of model fit test IRT 1 PL, 2 PL. and 3 PL with response data of each item of test question package A

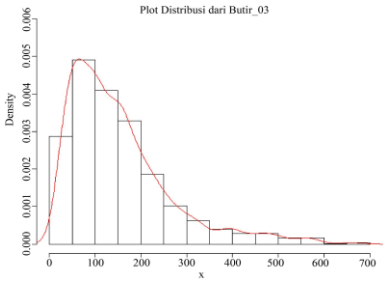
Test Item No.	1 PL	2 PL	3 PL
Item 01	Not suitable	Not suitable	Not suitable
Item 02	Not suitable	Not suitable	Not suitable
Item 03	Not suitable	Not suitable	Not suitable
Item 04	Not suitable	Not suitable	Not suitable
Item 05	Not suitable	Suitable	Suitable
Item 06	Not suitable	Suitable	Not suitable
Item 07	Not suitable	Suitable	Suitable
Item 08	Not suitable	Suitable	Suitable
Item 09	Not suitable	Not suitable	Not suitable
Item 10	Not suitable	Not suitable	Not suitable
Item 11	Not suitable	Not suitable	Not suitable
Item 12	Not suitable	Suitable	Suitable
Item 13	Not suitable	Not suitable	Not suitable
Item 14	Not suitable	Not suitable	Not suitable
Item 15	Not suitable	Suitable	Suitable



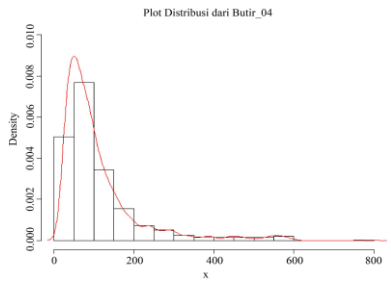
Distribution Plot from Item 01



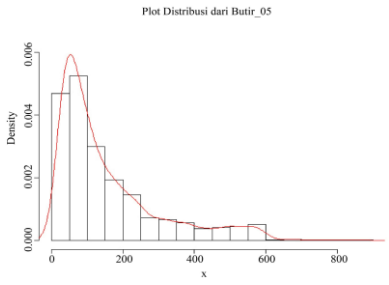
Distribution Plot from Item 02



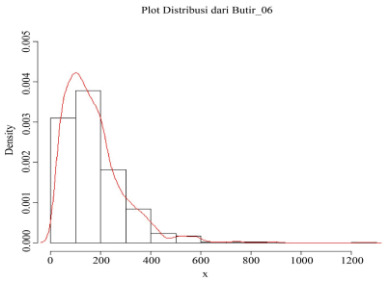
Distribution Plot from Item 03



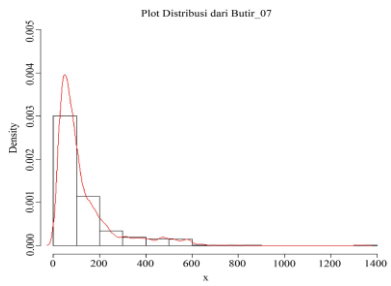
Distribution Plot from Item 04



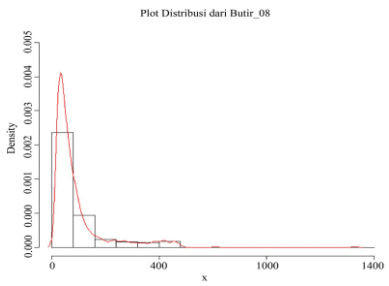
Distribution Plot from Item 05



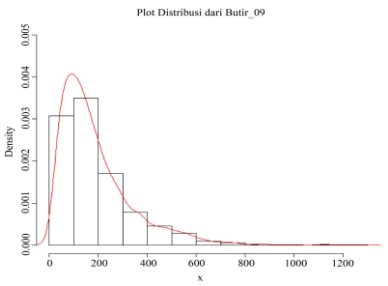
Distribution Plot from Item 06



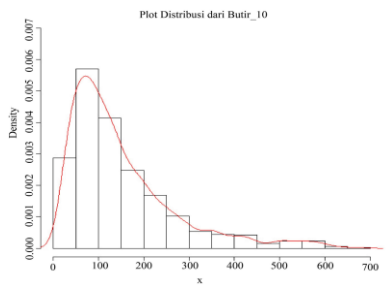
Distribution Plot from Item 07



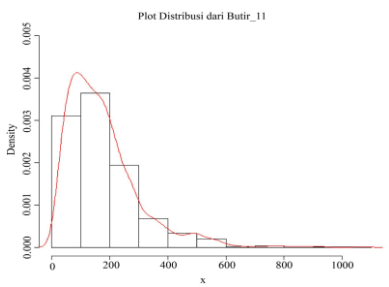
Distribution Plot from Item 08



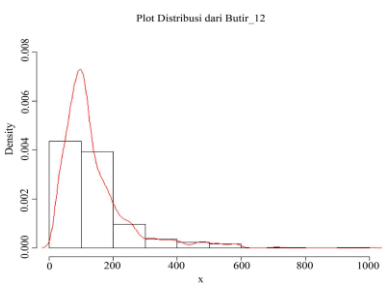
Distribution Plot from Item 09



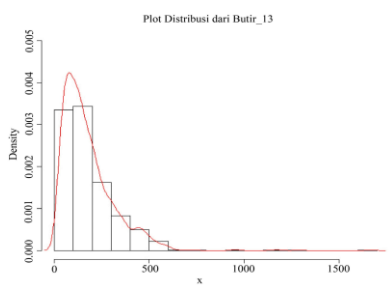
Distribution Plot from Item 10



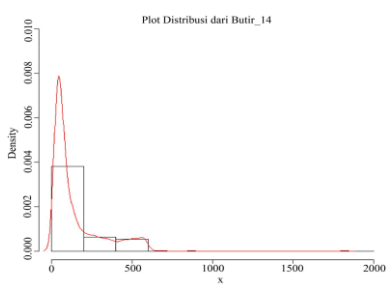
Distribution Plot from Item 11



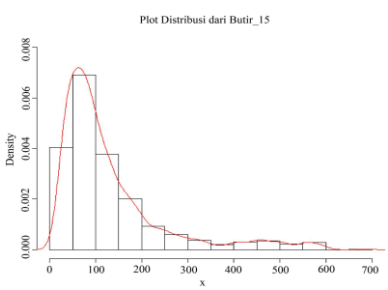
Distribution Plot from Item 12



Distribution Plot from Item 13



Distribution Plot from Item 14



Distribution Plot from Item 15

Fig. 2 Data distribution plot of response time

Based on Figure 2 above, we can describe that the response time data for the 15 items above is positive with a low arithmetic mean and high variance; the data distribution is positively skewed, which indicates that the probability of test takers answering quickly (short response time) is greater than the probability of test takers answering with a long time.

Researchers use the Lognormal distribution model to model response time data. Lognormal distributed response time data if transformed using logarithmic, will spread following the Normal distribution [20]. Response time random variables that are Lognormal distributed when algorithmized will spread following the normal distribution, so:

$$\ln t_{ij} \approx N\left(\left(\beta_j - \tau_i\right), \sigma^2\right) \quad (4)$$

equation as follows:

$$f\left(t_{ij}|\beta.\tau\right) = \frac{1}{t_{ij}\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij}-(\beta_j-\tau_i))}{\sigma}\right]^2\right) \text{ for t > 0} \quad (5)$$

$\left(t_{ij}\right)$ is the response time of the i-th test taker on the j-th question. Estimation of the Lognormal model in the equation using the Bayesian method produces parameter estimates of the speed of the i-th test taker answering the question (τ_i)question item$\left(\beta_j\right)$.

3.1.2. Second Level Modeling

The modeling structure at the second level is composed of a joint model of estimating person parameters and item parameters that have been estimated at the first level of modeling. Person parameters, namely the test taker ability parameter (θ_i) and the speed parameter of the i-th test taker answering the question (τ_i)normal distribution model is modeled using a bivariate, with the following equation:

$$f(\theta.\tau|\mu.\sigma.\rho_{\theta\tau}) = \frac{1}{\sigma_{\boldsymbol{\theta}}\sigma_\tau 2\pi\sqrt{1-\rho_{\theta\tau}^2}} exp\left\{-\left(\frac{1}{(2\sqrt{1-\rho_{\theta\tau}^2})}\right)\left[\frac{(\boldsymbol{\theta_i}-\mu_{\boldsymbol{\theta}})^2}{\sigma_{\boldsymbol{\theta}}^2} - \frac{2\rho_{\theta\tau}(\boldsymbol{\theta_i}-\mu_{\boldsymbol{\theta}})(\boldsymbol{\tau_i}-\mu_\tau)}{\sigma_{\boldsymbol{\theta}}\sigma_\tau} + \frac{(\boldsymbol{\tau_i}-\mu_\tau)^2}{\sigma_\tau^2}\right]\right\} \quad (6)$$

μ_θ and μ_τ is the average value θ_i and τ_i; while $\sigma_{\boldsymbol{\theta}}$ and σ_τ is the standard deviation θ_i and τ_i; and $\rho_{\theta\tau}$ is the person rho coefficient between (θ_i) and (τ_i). To make it easier to write, the equation is written in matrix form with the equation as below:

$$f(\xi_i|\mu_i,\Sigma_i) = \frac{|\Sigma_i^{-1}|^{1/2}}{2\pi} exp\left[-\frac{1}{2}(\xi_i-\mu_i)^T\Sigma_i^{-1}(\xi_i-\mu_i)\right] \quad (7)$$

$$\text{with } \xi_i = (\theta_i,\tau_i) \text{ and } \mu_i = (\mu_\theta,\mu_\tau)^T \text{ and } \Sigma_i = \begin{pmatrix}\sigma_\theta^2 & \sigma_{\theta\tau}\\ \sigma_{\theta\tau} & \sigma_\tau^2\end{pmatrix}$$

Item parameters, namely test item difficulty and test item time intensity, are modeled together using the bivariate normal distribution model equation, with the following equation:

$$f\left(b.\beta|\mu.\sigma.\rho_{b\beta}\right) = \frac{1}{\sigma_b\sigma_\beta 2\pi\sqrt{1-\rho_{b\beta}^2}} exp\left\{-\left(\frac{1}{(2\sqrt{1-\rho_{b\beta}^2})}\right)\left[\frac{(b_j-\mu_b)^2}{\sigma_b^2} - \frac{2\rho_{b\beta}(b_j-\mu_b)(\beta_j-\mu_\beta)}{\sigma_b\sigma_\beta} + \frac{(\beta_j-\mu_\beta)^2}{\sigma_\beta^2}\right]\right\} \quad (8)$$

μ_b and $\mu_{\boldsymbol{\beta}}$ is the average value b_j and β_j; while σ_b dan $\sigma_{\boldsymbol{\beta}}$ is the standard deviation b_j and β_j; and $\rho_{b\boldsymbol{\beta}}$ is the item correlation coefficient (item rho) between the item difficulty level b_j and item time intensity β_j. To make it easier to write, equation (8) is written in matrix form with the equation as below:

$$f(\psi_j;\mu_j,\Sigma_j) = \frac{\left|\Sigma_j^{-1}\right|^{1/2}}{(2\pi)^{2/2}} exp\left[-\frac{1}{2}\left(\psi_j-\mu_j\right)^T\Sigma_j^{-1}\left(\psi_j-\mu_j\right)\right] \quad (9)$$

$$\text{With } \psi_j = (b_j,\beta_j) \text{ with } \mu_j = \left(\mu_b,\mu_\beta\right) \text{ and } \Sigma_j = \begin{pmatrix}\sigma_b^2 & \sigma_{b\beta}\\ \sigma_{b\beta} & \sigma_\beta^2\end{pmatrix}.$$

To estimate the parameters of the Bivariate Normal model in equations (8) and (9) above, we use the Bayesian method. The steps of parameter estimation using Bayesian are as follows:

1. Determine the likelihood distribution that contains the empirical data information.

Since independent of t_{ij} then the likelihood distribution function $\left(y_{ij}.t_{ij}\right)$ with i=1. n and j=1.2…. k is as follows:

$$L\left(y_{ij}.t_{ij}|\xi_i.\psi_j\right) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left\{f\left(y_{ij}|\theta_i.a.b_j\right)f\left(t_{ij}|\tau_i.\alpha_j.\beta_j\right)f(\xi_i|\mu_i.\Sigma_i)f(\psi_j|\mu_j.\Sigma_j)\right.$$

$$f(y_{ij}.t_{ij}|\xi_i.\psi_j) = \prod_{i=1}^{n}\prod_{j=1}^{k}\left\{\left(\left(\frac{\exp\left(a_j(\theta_i-b_j)\right)}{1+\exp\left(a_j(\theta_i-b_j)\right)}\right)^{y_{ij}}\left(1-\left(\frac{\exp\left(a_j(\theta_i-b_j)\right)}{1+\exp\left(a_j(\theta_i-b_j)\right)}\right)\right)^{1-y_{ij}}\right)x$$
$$\left(\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\left(-\frac{1}{2}\left[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))\right]^2\right)\right)x\left(\frac{\left|\Sigma_j^{-1}\right|^{\frac{1}{2}}}{(2\pi)^2}exp\left[-\frac{1}{2}(\psi_j-\mu_j)^T\Sigma_j^{-1}(\psi_j-\mu_j)\right]\right)x$$
$$\left(\frac{\left|\Sigma_i^{-1}\right|^{1/2}}{2\pi}exp\left[-\frac{1}{2}(\xi_i-\mu_i)^T\Sigma_i^{-1}(\xi_i-\mu_i)\right]\right)\Bigg\} \quad (10)$$

2. Determining the Prior Distribution

The Prior distribution used for the estimation of covariance variance Σ is the Inverse-Wishart normal distribution (Box & Tio, 1973 and Gelman, Carlin, Stearn & Hall, 1995).

$$f(\mu_j.\ \Sigma_j) = f(\Sigma_j)f(\mu_j|\Sigma_j)$$
$$f(\mu_j,\Sigma_j) = |\Sigma_{j0}|^{-\left(\frac{(v_{j0}+3)}{2+1}\right)}exp\left(-\frac{1}{2}tr(\Sigma_{j0}\Sigma_j^{-1}) - \frac{k_{j0}}{2}(\mu_j-\mu_{j0})^T\Sigma_j^{-1}(\mu_j-\mu_{j0})\right)$$

3. Determine the Joint Posterior Distribution.

$$f(\xi,\psi,\mu_i,\mu_j,\Sigma_i,\Sigma_j|y,t) \propto \prod_{i=1}^{n}\prod_{j=1}^{k}f(y_{ij}|\theta_i,a_j,b_j)f(t_{ij}|\tau_i,\alpha_j,\beta_j)$$
$$f(\xi_i|\mu_i,\sum i)f(\psi_j|\mu_j,\sum_j)f(\mu_i,\sum i)f(\mu_j,\sum_j) \quad (11)$$
$$f(\xi_i.\psi_j.\mu_i.\mu_j.\ \Sigma_i.\ \Sigma_j|y_{ij}.t_{ij}) \propto \prod_{i=1}^{n}\prod_{j=1}^{k}\left\{\left(\frac{\exp\left(a_j(\theta_i-b_j)\right)}{1+\exp\left(a_j(\theta_i-b_j)\right)}\right)^{y_{ij}}\left(1-\left(\frac{\exp\left(a_j(\theta_i-b_j)\right)}{1+\exp\left(a_j(\theta_i-b_j)\right)}\right)\right)^{1-y_{ij}}\right.$$
$$\left(\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\left(-\frac{1}{2}\left[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))\right]^2\right)\right)x$$
$$\left(\frac{\left|\Sigma_j^{-1}\right|^{\frac{1}{2}}}{(2\pi)^2}exp\left[-\frac{1}{2}(\psi_j-\mu_j)^T\Sigma_j^{-1}(\psi_j-\mu_j)\right]\right)x$$
$$\left(\frac{|\Sigma_i^{-1}|^{1/2}}{2\pi}exp\left[-\frac{1}{2}(\xi_i-\mu_i)^T\Sigma_i^{-1}(\xi_i-\mu_i)\right]\right)$$
$$\left(|\Sigma_{i0}|^{-\left(\frac{(v_{i0}+3)}{2+1}\right)}\exp\left(-\frac{1}{2}tr(\Sigma_{i0}\Sigma_i^{-1}) - \frac{k_{i0}}{2}(\mu_i-\mu_{i0})^T\Sigma_i^{-1}(\mu_i-\mu_{i0})\right)\right)x$$
$$\left.\left(|\Sigma_{i0}|^{-\left(\frac{(v_{i0}+3)}{2+1}\right)}\exp\left(-\frac{1}{2}tr(\Sigma_{i0}\Sigma_i^{-1}) - \frac{k_{i0}}{2}(\mu_i-\mu_{i0})^T\Sigma_i^{-1}(\mu_i-\mu_{i0})\right)\right)\right\}$$

The value of the model parameters is determined through the expectation value (mean) of the random variable from the conditional posterior distribution function for the person in ordinary analytics, so it requires the help of the parameters that are difficult to obtain Gibbs sampling algorithm with the help of WinBUGS and R2WinBUGS software.

3.2. Comparison of Accuracy Parameter Estimation Results between Joint Model and Separate Model on Empirical Data

To find out the accuracy of the parameters of the joint model Two Parameter Logistik (2 PL) and response time models and separated models, the two models were implemented on empirical data to find the effect of response time on the accuracy of the parameters of the two models. The empirical data in this study is in the form of student answer responses and answer time records of 1559 students on 15 Mathematics item tests.

Model parameters with smaller standard deviation values indicate that the parameter estimators in the model are considered more accurate when compared to parameter estimators with larger standard deviation values. The process of estimating the parameters of the joint model and the separate model was carried out with the help of R2WinBUGS software. Model parameters were estimated using Bayesian MCMC Gibbs Sampling with 20,000 iterations, and the first burnin until the 2500th iteration burnin was ignored. With 20,000 iterations, it shows that the estimation process has reached a convergent condition. The parameter estimation process in the joint model with hierarchical response time on empirical data takes approximately 1486 seconds, while parameter estimation in the separated model with response time takes approximately 1178 seconds. o compare the effect of response time on the accuracy of the measurement results of test taker ability parameters and test item characteristics, the following is a comparison of the accuracy of parameter estimation results in the joint model with the separate model.

Table 3. Parameter estimation results of item time intensity and standard deviation in joint model and separate model

Parameters	Joint		Separate	
	Parameters	Std. Deviation	Parameters	Std. Deviation
Beta[1]	5,073	0,01585	5,027	0,02941
Beta[2]	4,560	0,01931	4,512	0,03169
Beta[3]	4,774	0,01914	4,726	0,03179
Beta[4]	4,429	0,0182	4,381	0,03101
Beta[5]	4,66	0,02302	4,611	0,03421
Beta[6]	4,913	0,01789	4,866	0,03041
Beta[7]	4,492	0,02134	4,443	0,03303
Beta[8]	4,359	0,02159	4,311	0,03332
Beta[9]	4,967	0,019	4,92	0,03089
Beta[10]	4,747	0,01893	4,699	0,03152
Beta[11]	4,933	0,018	4,886	0,03066
Beta[12]	4,708	0,0172	4,66	0,03044
Beta[13]	4,912	0,01919	4,865	0,03151
Beta[14]	4,495	0,02516	4,446	0,03576
Beta[15]	4,575	0,01926	4,527	0,0317
Average	4,721	0,0184	4,638	0,0296

Table 4. Estimation results and standard deviation of item difficulty level parameters (b) in the joint model and separate model

Parameters	Joint		Separate	
	Parameters	Std. Deviation	Parameters	Std. Deviation
b[1]	-0,2602	0,05473	-0,2603	0,05807
b[2]	-2,058	0,07977	-2,069	0,08004
b[3]	-0,8033	0,0822	-0,8051	0,07694
b[4]	-2,055	0,08029	-2,057	0,0813
b[5]	-0,8805	0,05575	-0,8792	0,0561
b[6]	-1,06	0,0583	-1,064	0,05849
b[7]	-1,533	0,07074	-1,554	0,07793
b[8]	-0,8963	0,0559	-0,8972	0,05583
b[9]	-1,288	0,06469	-1,34	0,07447
b[10]	-0,6873	0,05441	-0,6931	0,05507
b[11]	-0,662	0,05504	-0,6632	0,0565
b[12]	-1,72	0,07502	-1,773	0,07692
b[13]	-0,651	0,0587	-0,6537	0,05964
b[14]	-1,545	0,06628	-1,546	0,06623
b[15]	-1,464	0,06572	-1,469	0,06813
Average	-1,17091	0,065169	-1,18159	0,066777

Beta is the item time intensity parameter, which is the ideal time required by the item for the item to be answered. Table 3 shows that the joint model of IRT 2 PL with the response time model produces an estimated value of the item time intensity parameter (Beta) located in the interval 4.575 to 5.074 with an average of 4.721 which, when converted in seconds, will be equal to 282, 6 seconds. This shows that the average ideal time taken by each item to be answered is 283.26 seconds. While in the separated model, the estimated value lies in the interval 4.311 to 5.027 with an average of 4.638, which is equivalent to 278.28 seconds lower than the average IRT 2 PL model involving response time. If an item has a high or complex level of difficulty, it requires a deeper cognitive process, so the item time-intensity (Beta) required by the item is also high. The average value of the standard deviation of the item intensity parameter in the joint model of IRT 2 PL with a response time of 0.0184 is smaller than the standard deviation of the separated model estimator with a response time of 0.0296.

Thus, the ability measurement model involving response time information can improve the accuracy of estimating item time intensity parameters when compared to measurement models that do not involve response time information. Table 4 shows that the joint model of IRT 2 PL with the response time model produces an estimated value of the item difficulty parameter (b) located in the interval -2.2602 to -2.058. While the separate model produces an estimated value of the item difficulty parameter in the interval -2.2603 to -2.069. The average value of the item difficulty parameter estimates in the joint IRT 2 PL model and the hierarchical response time of -1.17091 is greater than the separate model of -1.1816. While the average standard deviation of the item difficulty parameter in the joint model of IRT 2 PL and response time model is 0.0651, smaller than the separate model of 0.0668. Thus, involving response time information in the ability measurement model is proven to improve the accuracy of the results of estimating item difficulty parameters because the standard deviation value of the parameters is smaller.

Table 5. Estimation results and standard deviation of item distinctiveness parameters (a) Joint model and separate model

Parameters	Hierarchy		Separate	
	Parameters	Std. Deviation	Parameters	Std. Deviation
a[1]	0,9753	0,1366	0,7156	0,2264
a[2]	0,05214	0,04095	0,04413	0,05072
a[3]	2,806	0,1823	0,9885	0,1735
a[4]	0,2433	0,09875	0,1523	0,1592
a[5]	0,1732	0,08073	0,135	0,1159
a[6]	0,2772	0,0832	0,1396	0,1418
a[7]	0,7741	0,1321	0,5344	0,2089
a[8]	0,09279	0,04772	0,05818	0,0769
a[9]	0,606	0,1291	0,5958	0,193
a[10]	0,28	0,09781	0,288	0,1384
a[11]	0,5233	0,1042	0,378	0,1627
a[12]	0,553	0,1234	0,3127	0,1976
a[13]	0,961	0,116	0,5857	0,2225
a[14]	0,05687	0,04202	0,04681	0,05406
a[15]	0,3589	0,1127	0,2652	0,1673
Average	0,5822	0,1018	0,3492	0,1525

Table 6. Estimation results and standard deviation of the ability (THETA) and speed (TAU) parameters of 1559 test takers on the joint model and separate model

Parameters	Hierarchy		Separate	
	Parameters	Std. Deviation	Parameters	Std. Deviation
Theta	0,054629	0,05391841	0,014973	0,284226
Tau	0.096216	0,13381911	-0,04613	0,185526

Table 5 describes that the joint model of IRT 2 PL and the response time model produces the value of the parameter estimator of item differentiation (a) located in the interval 0.0521 to 2.8060. While the separate model produces the estimated value of the item difficulty parameter in the interval 0.0441 to 0.9885. The average value of the item difficulty parameter estimator in the joint model of 0.5822 is greater than the separate model of 0.3492. While the average standard deviation of the item difficulty parameter in the joint model of 0.1018 is smaller than the separate model of 0.1525. Thus, involving response time information in the ability measurement model is proven to improve the accuracy of the results of estimating item difficulty parameters because the standard deviation value of the parameters is smaller. Table 6 shows that the average test taker ability parameter estimate (Theta) in the joint model of IRT 2 PL and response time model of 0.0546 is greater than the separate model of 0.0149. The average parameter estimates of test takers' answering speed (Tau) in the joint model of 0.0962 is greater than the separate model of -0.0461. While the average standard deviation of theta parameter estimates in the joint model of 0.0539 is relatively smaller than the standard deviation of the separate model of 0.2842, as well as the average standard deviation of Tau parameter estimates in the joint model of 0.1338 is relatively smaller than the standard deviation of the separate model of 0.1855.

Thus, models involving response time can improve the accuracy of estimating test-taker ability parameters when compared to parameters in separate models.

From the results of the comparative analysis of person parameters (ability and speed) and item parameters (item time intensity, item difficulty and item differentiability) show that the joint model produces more accurate parameters, more thorough than the separate model.

Therefore, the researcher concluded that the IRT 2 PL measurement model involving response time in this study is suitable for implementation in tests that are limited by time.

One of the advantages of the joint model of IRT 2 PL and the response time model is that the model can explain the dependency or correlation between person parameters (person rho) and the correlation between item parameters (item rho).

Person Rho describes the dependency relationship between the test taker's ability parameter (Theta) and the test taker's answering speed (Tau). While item rho describes the relationship between the item difficulty parameter (b) and the item time intensity parameter (Beta). The coefficients of person rho and item rho are as follows:

Table 7. Correlation coefficient of ability and speed parameters (Person Rho) and correlation coefficient of item time intensity and item difficulty parameters (Item Rho)

Type of Approach		*Person Rho* ($\rho_{\theta\tau}$)	*Item Rho* ($\rho_{b\beta}$)
Joint Model	Parameters	0,6482	0,7338
	Std. Deviation	0,0397	0,1191
Separate Model	Parameters	-	-
	Std. Deviation	-	-

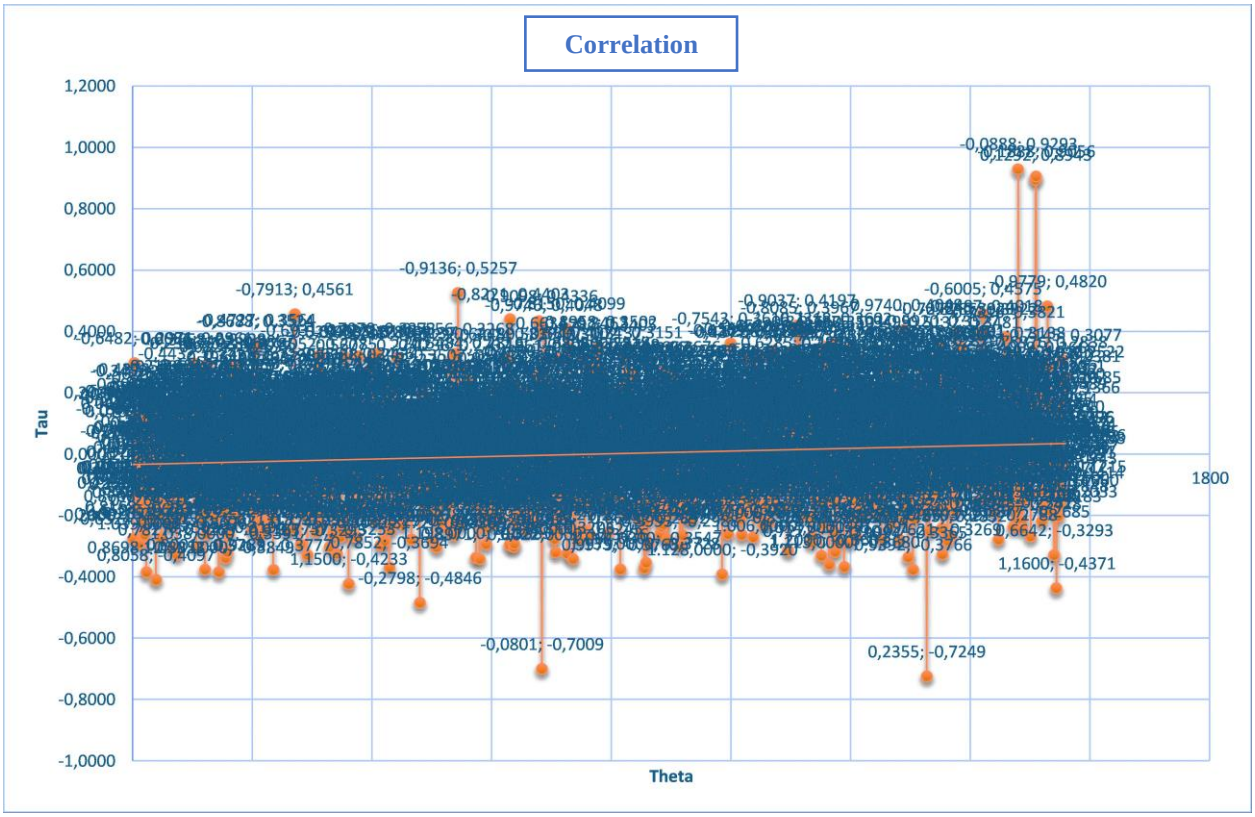


Fig. 3 Pattern of relationship between ability parameters (Theta) with answering speed parameters (Tau) 1559 new student admission test participants

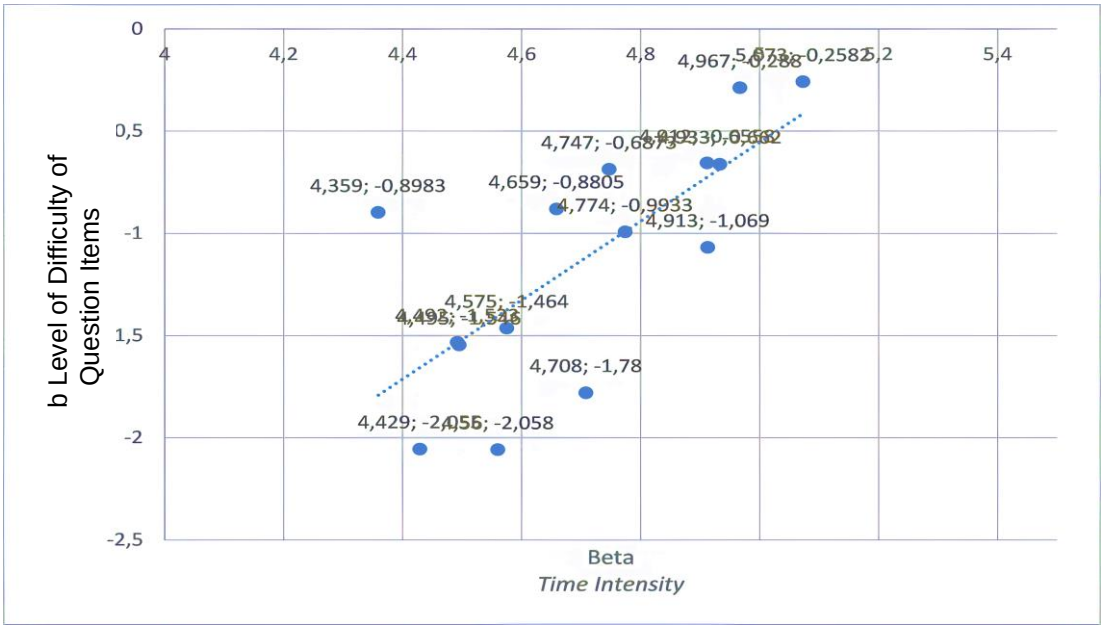


Fig. 4 Pattern of relationship between item time intensity (Beta) with item difficulty level (b) parameter 1559 participants of the 2019 new student admission selection test

The relationship pattern between the ability and speed parameters can be seen in Figure 3. From Figure 3, we can notice that there is a positive relationship pattern between the test taker's ability parameter and the speed of answering test items of 0.6482. This shows that the increase in the ability of test takers to answer items is directly proportional to the increase in the speed of answering items. From this, it can be concluded that there is a tendency for students with higher abilities to answer items more quickly, and students with lower abilities tend to answer items more slowly. The pattern of the relationship between the ability parameters of item difficulty and item intensity can be seen in Figure 4.

Figure 4 informs about the positive relationship pattern between the item difficulty parameter and the item intensity parameter of 0.7338. This positive relationship pattern informs us that the increase in item difficulty is directly proportional to the intensity of the time required for the item to be answered. This shows that the more the level of item difficulty increases, the ideal time needed for the test item to be answered also longer because difficult items require a deeper, more intense cognitive process to understand the meaning of the question or problem on the item and the process of finding problem-solving to answer the item.

To choose a suitable model (fit model) between the joint model of IRT 2 PL with response time (joint model) in a hierarchical manner or a separate model on empirical data, the researcher uses the Deviance Information Criterion (DIC) statistic. The DIC value in the hierarchical joint model will be compared with the DIC value in the RT 2 PL separate model with response time (separated model). Good fit in complex models is characterized by small DIC values [21]. A model with a small DIC indicates that the model is a good fit with the empirical data. The results of the comparison of DIC values between the two models are as follows:

Table 8. Empirical data DIC values in the 2 PL IRT joint model with hierarchical response time by separate modeling

Model	Together	Separate
Dbar	76. 417. 600	76.540,800
Dhat	75. 033. 800	74. 442, 400
pD	1.383, 860	2. 098, 390
DIC	77. 801, 500	78.639, 200

Table 8 provides information that the joint model of IRT 2 PL with the response time model shows a DIC statistical value that is smaller than the DIC value of IRT 2 PL separated by response time (separated model). This case proves that the IRT 2 PL model that involves response time (joint model) in the measurement process is more suitable (fit) with empirical data in describing the real conditions of computer-based testing when compared to IRT 2 PL separated with response time (separated model).

4. Conclusion

Based on the results of the data and discussion, several conclusions can be taken, such as:

(1) The test taker ability measurement model developed in this study is in the form the joint model of IRT 2 PL with Lognormal model produces the accuracy of estimating parameters when compared to parameters in a separate model.
(2) This joint model is considered capable of explaining the real conditions of computer-based testing with limited answer time.
(3) This joint model uses response accuracy data as an indicator measuring the ability of test takers and response time data as an indicator measuring the speed of answering test items.
(4) The joint model of IRT 2 PL and the response time model are more accurate than the separate model.
(5) The DIC value in the joint model of IRT 2 PL with the response time model is smaller than the separate model, so the joint model is more suitable (model fit) implemented on empirical data. From the joint model, it can be identified that there is a positive correlation of 0.6482 between the ability and speed parameters and a positive correlation of 0.7338 between the Item Time Intensity and Item Difficulty Level Parameter.

Acknowledgment

We would like to express our deepest gratitude and appreciation to Universitas Negeri Yogyakarta (UNY) for providing the necessary resources and facilities that supported the completion of this research. We extend our sincere thanks to our research advisors and mentors for their invaluable guidance, expertise, and continuous support throughout the research process. Their constructive feedback and encouragement have significantly contributed to shaping the direction and quality of this study.

References

[1] Ariel Rubinstein, "Response Time and Decision Making: An Experimental Study," *Judgment and Decision Making*, vol. 8, no. 5, pp. 540-551, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[2] Wim J. Van Der Linden, "A Hierarchical Framework for Modeling Speed and Accuracy on Test Items," *Psychometrika*, vol. 72, pp. 287-308, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[3] Paul De Boeck, and Minjeong Jeon, "An Overview of Models for Response Times and Processes in Cognitive Tests," *Frontiers in Psychology*, vol. 10, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yulia A. Dodonova, and Yury S. Dodonov, "Faster on Easy Items, More Accurate on Difficult Ones: Cognitive Ability and Performance on a Task of Varying Difficulty," *Intelligence*, vol. 41, no. 1, pp. 1-10, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[5] Jean-Paul Fox, and Sukaesi Marianti, "Person-Fit Statistics for Joint Models for Accuracy and Speed," *Journal of Educational Measurement*, vol. 54, no. 2, pp. 243-262, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ronald K. Hambleton, and Hariharan Swaminathan, *Item Response Theory Principles and Applications*, Kluwer-Nijhoff Pub, pp. 1-332, 1989. [Google Scholar] [Publisher Link]

[7] Ronald K. Hambleton, Hariharan Swaminathan, and H. Jane Rogers, *Fundamentals of Item Response Theory*, SAGE Publications, pp. 1-174, 1991. [Google Scholar] [Publisher Link]

[8] Gerard J.P. Van Breukelen, "Psychometric Modeling of Response Speed and Accuracy with Mixed and Conditional Regression," *Psychometrika*, vol. 70, pp. 359-376, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[9] Wim J. Van Der Linden, and Ronald K. Hambleton, *Item Response Theory*: *Brief History*, *Common Models*, *and Extensions*, Handbook of Modern Item Response Theory, pp. 1-28, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yi-Hsuan Lee, and Haiwen Chen, "A Review of Recent Response-Time Analyses in Educational Testing," *Psychological Test and Assessment Modeling*, vol. 53, no. 3, pp. 359-379, 2011. [Google Scholar] [Publisher Link]

[11] D. Thissen, "Latent Trait Scoring of Timed Ability Tests," *Computerized Adaptive Testing Conference*, Wayzata, MN, USA, pp. 1-467, 1979. [Google Scholar] [Publisher Link]

[12] Wim J. Van Der Linden, "Conceptual Issues in Response-Time Modeling," *Journal of Educational Measurement*, vol. 46, no. 3, pp. 247-272, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[13] Andrew Gelman et al., *Bayesian Data Analysis*, 1st ed., Chapman and Hall/CRC, pp. 1-552, 1995. [CrossRef] [Google Scholar] [Publisher Link]

[14] David W. Hosmer, Jr. Stanley Lemeshow, and Rodney X. Sturdivant, *Applied Logistic Regression*, Wiley, pp. 1-528, 2013. [Google Scholar] [Publisher Link]

[15] Prathiba Natesan, "A Review of Bayesian Item Response Modeling: Theory and Applications," *Journal of Educational and Behavioral Statistics*, vol. 36, no. 4, pp. 550-552, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[16] Ioannis Ntzoufras, *Bayesian Modeling Using WinBUGS*, Wiley, pp. 1-520, 2011. [Google Scholar] [Publisher Link]

[17] J.K. Lindsey, *Statistical Analysis of Stochastic Processes in Time*, Cambridge University Press, pp. 1-338, 2004. [Google Scholar] [Publisher Link]

[18] Heri Retnawati, *Item Response Theory and its Applications: For Researchers, Measurement and Testing Practitioners, Graduate Students*, Yogyakarta, Nuha Medika, pp. 1-200, 2014. [Google Scholar] [Publisher Link]

[19] Frederic M. Lord, Melvin R. Novick, and Allan Birnbaum, *Statistical Theories of Mental Test Scores*, Information Age Pub, pp. 1-568, 2008. [Google Scholar] [Publisher Link]

[20] Deborah L. Schnipke, and David J. Scrams, *Exploring Issues of Examinee Behavior: Insights Gained from Response-Time Analyses*, 1st ed., Computer-Based Testing, pp. 1-30, 2002. [Google Scholar] [Publisher Link]

[21] T. Loeys, Yves Rosseel, and Kristof Baten, "A Joint Modeling Approach for Reaction Time and Accuracy in Psycholinguistic Experiments," *Psychometrika*, vol. 76, pp. 487-503, 2011. [CrossRef] [Google Scholar] [Publisher Link]