

## Two Parameter Logistic Model with Lognormal Response Time for Computer-Based Testing

<https://doi.org/10.3991/ijet.v14i15.10580>

Andhita Dessy Wulansari <sup>(✉)</sup>

Universitas Negeri Yogyakarta, Sleman, Indonesia  
Institut Agama Islam Negeri Ponorogo, Ponorogo, Indonesia  
andhita.dessy2015@student.uny.ac.id

Kumaidi

Universitas Muhammadiyah Surakarta, Sukoharjo, Indonesia

Samsul Hadi

Universitas Negeri Yogyakarta, Sleman, Indonesia

**Abstract**—In addition to the information on response pattern/accuracy, Computer-Based Testing (CBT) can also generate information on response time. This research contributes to develop two-parameter logistic model with random variables of response time for CBT power test. Based on that model, the probability of the test taker answering the test questions correctly is influenced by the test taker's ability, the test question's discriminating power, the question's level of difficulty, the delay due to the test question's factors and response time. The development of this model aims to improve the parameter estimation of the logistic model on Item Response Theory (IRT) which does not consider the response time of the model. This model is simultaneously developed using joint distribution concept, by multiplying the conditional distribution of response accuracy (two-parameter logistic model) by response time with marginal distribution of response time. The marginal distribution chosen in this study is lognormal distribution because it has positive value in the form of positive skewed according to the characteristics of the response time. To prove the model is suitable for power test, is tested using CBT data. The study found that the simultaneous model generated from the multiplication between the two-parameter logistic model integrated with response time and the lognormal response time model is an appropriate model for CBT power test.

**Keywords**—Logistic Model, Response Time, Lognormal, Joint Distribution, CBT.

### 1 Introduction

Due to the advancement of technology and information, conventional tests using papers or Paper Based Test (PBT) has been replaced with tests using computers or Computer Based Testing (CBT). By conducting tests using computers, complex in-

formation not only about the response patterns from the test takers but also about the response time on how they answer each of the questions can be obtained [1]. Response time can be described as time required by a test taker to read and answer one question on the test [2]. In understanding the behaviour of the test takers, response time is one of the most important variables considered during the test. Response time can be used to identify the type of strategy used by the test participants, whether they just guess the answer or they do read and answer the questions carefully. This guessing behaviour is probably done by test takers to answer the questions because they realize that they will run out of time soon [3].

Regardless of the PBT or CBT method, there are two main parameters used in a test. The first parameter is response accuracy that shows how true the test takers answer the question, and the second one is response time that shows how fast the test takers answering the question although not all test organizers record the information on response time. Using these two parameters, the information on whether the test participants pass or fail the test and whether they are categorized into slow or fast can be obtained [4]. Therefore, in the measurement process, response time should also be considered because it can improve the parameter estimation on IRT (Item Response Theory) model that only consider response accuracy. This statement is supported by Schnipke and Scrams [5] who also consider the response time on their models to estimate the ability of the test takers. Hornke [6] also argues that response time is an essential variable that should be taken into account as it shows the behaviour of the test takers that has to be distinguished from the score in answering the questions. Abdelfattah [7] showed in his study that there was a difference in the parameter estimation of the test takers' ability and the parameter of test questions between the models that consider and do not consider response time.

According to Ethink [8], mathematically response time models can be done in three ways:

- Exclusive response time model, that is done if the questions tested are easy and the time given is very limited
- Separate response time model, in which the information on the response time and the accuracy of the score are obtained separately/ independently
- Simultaneous response time model, that is obtained using the joint distribution concept by combining accuracy parameter and response time parameter into the logistic model using random variable of response time simultaneously. Simultaneous model is a complex mathematical model as it has many parameters. Bayesian method is considered appropriate to estimate the parameter of test takers and test questions. According to Natesan [9], Bayesian method can estimate parameter models that are very complex, violate IRT basic assumptions, and have very small sample size. The numerical solution to this complex estimation of parameter model can be done using Markov Chain Monte Carlo method [10].

In developing response time model, test types should also be taken into consideration, whether it is a speed test or power test. According to Gulliksen [11], pure speed test is a test done in limited time whose level of the difficulty of the questions is the same (easy to answer correctly) and then the score is calculated based on the total

amount of time spent by the test taker to answer all the questions. Meanwhile, pure power test is a test in limited time whose level of difficulty of the questions varies and the score is calculated based on the number of questions answered correctly. In practice, pure power test does not exist because every test is always conducted with time constriction [12]. By implementing the time construction, there is other aspects involved in answering the test questions, i.e. speed aspect. Test takers' speed can be measured from their response time spent answering each of the questions. This model is developed due to the time constraint on the power test. Rouder, Province, Morey, Gomez, and Heathcote [13] developed response time model in speed test to measure the test taker's cognitive ability called Lognormal Race Model. Finishing time (response time) is assumed to have lognormal distribution as the distribution is considered to be suitable with the response time characteristics. According to Lindsey [14], there are several characteristics worth considering in choosing the distribution of response time:

- It should be positive
- Positive skewed (the probability of short response time is bigger than that of long response time). The distributions considered suitable to both characteristics are lognormal distribution, weibull and gamma [14].

Each developed response time model has its own specifications, weaknesses and strengths. There is no standard model that can be used to model all response time cases in every test because the distribution of response time can be various [15]. Therefore, the development of model considering the response time is continuously done to improve the current existing model to get the most reliable model. Not every response time model developed for power test is proven effective for power test, for instance, Ingrisone II [16] that is intended to be used in power test, according to Hidayah and Kumaidi [17] is proven more suitable for speed test because each of the questions is easy to answer correctly in limited time. This study aim to develop two-parameter logistics model with random variable of response time for power test, whose chosen parameters are limited to the difference of the questions ( $a_j$ ), the difficulty level of the questions ( $b_j$ ), the ability of the test takers ( $\Theta_i$ ), the delay due to the questions ( $d_i$ ) and response time ( $t_{ij}$ ). This model is simultaneously developed using joint distribution concept with lognormal marginal distribution. The resulted parameter model is estimated using Bayesian Markov Chain Monte Carlo method with Gibbs Sampling algorithm. The test on the model suitability is done using CBT data. The method used to test the suitability is Posterior Predictive Model Checking (PPMC) by comparing the difference between the data from observation and the result of model prediction.

## 2 Literature Review

### 2.1 Item response theory (IRT) model

Modern test theory that is commonly known as Item Response Theory (IRT) is developed by measurement experts to minimize the weaknesses of classical theory. Looking at the number of parameters, there are some models known as IRT, which are one-parameter logistic model, two-parameters logistic model, and three-parameters logistic model [18]. The main parameter in IRT calculation is the ability of the test takers. The parameter symbol of test taker's ability is  $\Theta$  (theta). The value interval of  $\Theta$  is infinite because it ranges from interval  $-\infty$  to  $\infty$ , however the value of  $\Theta$  can be determined at a standard interval, between -4 and 4 [19].

On one-parameter logistic model, the level of difficulty of the questions ( $b$ ) is defined as a value on the scale of test taker's ability whose probability to answer the questions correctly is 0.5 [19]. On one-parameter logistic model, the test taker's probability in answering the question correctly can be formulated as follows.

$$p_i(X_{ij} = 1 | \theta_i, b_j) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)} \quad (1)$$

On two-parameter logistic model, it involves not only the parameter of the questions' level of difficulty ( $b$ ), but also the parameter of discriminating power ( $a$ ). Parameter of discriminating power ( $a$ ) aims to identify whether a question can differentiate whether a group is measurable or not based on the differences existing in the group. The test taker's probability in answering the questions correctly for two-parameter logistic model can be formulated using the following formula [19]:

$$p_i(X_{ij} = 1 | \theta_i, b_j) = \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} \quad (2)$$

Three-parameter logistic model involves not only the parameter of the questions' level of difficulty ( $b$ ), the discriminating power ( $a$ ), but also the parameter of pseudo guessing ( $c$ ). Pseudo guessing parameter is the test taker's probability to guess the answer correctly. The formula for three-parameter logistic model [19] is as follows.

$$p_i(X_{ij} = 1 | \theta_i, b_j) = c_j + (1 - c_j) \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} \quad (3)$$

According to Hambleton, Swaminathan and Rogers [19], big size of sample has to be used to make the estimation parameter stable. While to estimate the ability, Maximum Likelihood Estimation (MLE) can be used.

## 2.2 Response time model

Response time or response latency is time required by a test taker to read and answer one test question [2]. Although the theory of model topics has been widely discussed, the discussion is not as much as the theory of IRT model. Response time model can provide information on how fast a test taker answers a test question. Speed variable also indicates the types of strategy used by the test taker, whether he/she just guesses or do the test seriously like how it is supposed to be.

The theory of response time has been reviewed in three different ways

- Exclusive response time model, that is done if the questions tested are easy and the time given is very limited
- Separate response time model, in which the information on the response time and the accuracy of the score are obtained separately/independently
- Simultaneous response time model, that is obtained using the joint distribution concept by combining accuracy parameter and response time parameter into the logistics model using random variable of response time simultaneously.

The chosen way used to develop the model in this study is the third model, that is simultaneous model between response accuracy and response time. Besides, according to Schnipke and Scrams [5], the development of simultaneous model done by the experts are also based on three things, that are:

- Test type, whether it is speed test or power test
- The assumption of response time distribution used
- The assumption of the relation between response accuracy and response speed.

Research on the development of simultaneous model between response accuracy and response time has not been done much. Some experts who develop this are Roskam [20], Verhelst, Verstralen and Jansen [21], Wang and Hanson [22], Wang [23], Ingrisone [24], Ingrisone II [16] and Hidayah and Kumaidi [17].

Roskam [20], Verhelst, Verstralen and Jansen [21] involve response time parameter into Rasch model. However, both developed models have limitation as it can only be used in speed test. Roskam [20] implements weibull distribution to show the characteristics of response time, while Verhelst, Verstralen and Jansen [21] uses gamma distribution. Response time in both models is the total amount of time required by test takers to answer all test questions, not time required to answer one question. The following is model of Roskam [20]:

$$p(U_{ij} = 1 | t_i, j, i) = \frac{\theta_j t_{ij}}{\theta_j t_{ij} + \varepsilon_i} = \frac{\exp(\xi_j + \tau_{ij} - \sigma_i)}{1 + \exp(\xi_j + \tau_{ij} - \sigma_i)} \quad (4)$$

Based on the model above, the test taker's ability is symbolized as  $\theta$ , response time is symbolized as  $t$ , and the level of difficulty is symbolized as  $\varepsilon$ . The symbols  $\xi$ ,  $\tau$  and  $\sigma$  is the logarithm of  $\theta$ ,  $t$  and  $\varepsilon$ . Based on the relation between parame-

ters on the mentioned model above, it can be concluded that the test taker’s ability in answering the test questions correctly will rise if the time is given more, and the test taker’s speed is an important aspect influencing the time required by the test taker to answer one test question [12].

The limitation of Roskam [20], Verhelst, Verstralen and Jansen [21] is improved by Wang and Hanson [22] by implementing the parameter of response time into the three-parameter logistic model that, according to Wang and Hanson [22], is suitable for power test. In this model, response time is assumed as fixed variable, so it is considered to have four parameters that then is known as four-parameter logistic model.

Wang and Hanson [22] improve their model based on the assumption that response time is fixed variable in which there are many variables that may influence response time of a test taker (random variable), by multiplying four-parameter logistic model as conditional response distribution by time response using one-parameter weibull distribution as marginal response time (joint distribution concept). However, according to Wang [23], the implementation of one-parameter weibull distribution needs to be reviewed as it is considered too simple to read the reality in measurement practice. The following is model of wang [23]:

$$p_j(\theta_i = 1 | \theta_i, \rho_i, a_j, b_j, c_j, t_{ij}) = c_i + \frac{1 - c_i}{1 + \exp\left(-1.7a_j \left(\theta_i - \left(\frac{\rho_i d_j}{t_{ij}}\right) - b_j\right)\right)} \tag{5}$$

Based on the model above, a test taker’s ability is symbolized as  $\theta$ , discriminating power is symbolized as a, the level of difficulty is symbolized as b, and pseudo guessing is symbolized as c, while delay parameter due to test questions is symbolized as d, response time is symbolized as  $t$  and delay parameter due to the test taker is symbolized as  $\rho$  [23].

Ingrisone [24] and Ingrisone II [16] also develop a model using joint distribution concept. The differenci is, Ingrisone [24] multiplies two-parameter logistic model as conditional item response distribution by response time using modified lognormal marginal distribution, while Ingrisone II [16] multiplies one-parameter logistic model as conditional item response distribution by response time using two-parameter marginal weibul distribution. According to Ingrisone II [16], the developed simulataneous model is suitable for power test. However, in the next study, Hidayah and Kumaidi [17] shows that, in the same amount of response time and the different level of test takers’ ability, the probability of answering the test questions correctly is almost the same (speed test). Model of Ingrisone [24] is as follows.

$$p(U_{ij} = 1 | t_{ij}, \theta_i, \rho_i, a_j, b_j) = \frac{\exp(-1.7a_j(\theta_i - \eta t_{ij} - b_j))}{1 + \exp(-1.7a_j(\theta_i - \eta t_{ij} - b_j))} \tag{6}$$

Based on the model above, the test taker’s ability is symbolized as  $\theta$ , dicriminat-ing power is symbolized as a, the level of difficulty is symbolized as b, and response time is symbolized as  $t$ . Parameter  $\eta$  is coefficient regresion from time variable

whose value is constant.  $\eta$  will have value if the response time influences the probability of the test taker answering the test questions correctly, and  $\eta$  will be zero if the response time does not influence the probability of the test taker answering the test questions correctly Ingrisone [24]. The following is model of Ingrisone II [16].

$$p(X_{ij} = 1 | t_{ij}, \theta_i, b_j, \eta) = \frac{\exp\left(\eta(\theta_i - \eta t_{ij} - b_j)\right)}{1 + \exp\left(\eta(\theta_i - \eta t_{ij} - b_j)\right)} \quad (7)$$

Based on the model above, the test taker's ability is symbolized as  $\theta$ , the level of difficulty is symbolized as  $b$ , and response time is symbolized as  $t$ . Parameter  $\eta$  is coefficient regression from time variable whose value is constant.  $\eta$  will have value if the response time influences the probability of the test taker answering the test questions correctly, and  $\eta$  will be zero if the response time does not influence the probability of the test taker answering the test questions correctly, the mentioned model above will be one-parameter logistic model.

The limitation of Ingrisone II [16] is improved by Hidayah and Kumaidi [17] by implementing the parameter of response time into one-parameter logistic model using joint distribution concept for power test. Hidayah and Kumaidi [17] used marginal lognormal distribution in which  $\mu_{\ln t_{ij}}$  is defined as function of the speed and the effort required by a test taker in doing the test. According to Hidayah and Kumaidi [17], factors influencing both parameters need further reviewing. The following is model of Hidayah and Kumaidi [17]:

$$p(X = 1 | \theta_j, t_{ij}, b_i) = \frac{\exp\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)}{1 + \exp\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)} \quad (8)$$

Based on the model above, the test taker's ability is symbolized as  $\theta$ , the level of difficulty is symbolized as  $b$ , and response time is symbolized as  $t$ . If the value of  $t$  is infinite (test does not have time constraint), model of Hidayah and Kumaidi [17] is the same with one-parameter logistic model. The limitation of defining  $\mu_{\ln t_{ij}}$  in model of Hidayah and Kumaidi [17] can be used by other researchers as the basis of developing simultaneous model between response accuracy and response time to be a more realistic model.

### 2.3 Test

Test is a set of planned questions or tasks to obtain information on trait (educational or psychological attributes) in which each of the questions or tasks has a certain

answer regarded as correct [25]. Based on the definition, every test requires responses from a subject (test taker) that can be concluded as a trait that belongs to that person. In general, tests are usually used to measure the students' learning outcomes especially the cognitive outcomes related to the understanding of learning materials in accordance with the purpose of teaching and learning [26]. Tests are standard and objective measure that can be used in wide contexts to measure and compare the individual's psychological and behavioral state. Based on their purposes in educational contexts, tests can be divided into:

**Speed Test:** This test aims to evaluate the test taker's ability in terms of the speed of thinking or skill whether it is logical skill, rote skill, or the skill in understanding the subjects that have been learned. The time provided to answer all questions in the test is shorter compared to that in other tests, as the most important thing is the minimum time used to correctly answer as many questions as possible.

**Power Test:** This test aims to evaluate the test taker's ability in a certain field with no time constraint. The evaluated skills are both cognitive and psychomotor. The test questions are usually more difficult, involving various different concepts and it requires the test takers to implement problem solving and use all of their abilities in analyzing, synthesizing, and evaluating.

The test can be divided into two types, i.e. conventional and modern test. Conventional test is known as Paper Based Test (PBT) or Paper Pencil Test (PPT), while modern test is known as Computer-Based Testing (CBT). PBT or PPT uses paper and pencil with certain numbers of questions (fixed length) and the score calculation is done manually by human. Meanwhile, CBT uses computer with certain numbers of questions (fixed length) and the score calculation is done automatically using computer assistance.

The conventional test system is now regarded as less effective, efficient, and tends to cause problems so it is easily replaceable with modern test. In the advanced technological and informative era, it is very feasible to conduct a computer based test [27]. The existence of CBT makes the students easier to do self assessment as CBT is able to provide direct score and feedback. According to Linacre and Rudner, the advantages of CBT are the flexibility of the test management, the security of the test, the improved motivation to be IT literate, and the efficiency of time [28]. Table 1 explains the difference between PBT and CBT in more details.

**Table 1.** The Difference between PBT and CBT

PBT	CBT
The test is similarly designed for all test takers and it is done manually	The test is similarly designed for all test takers and it is done using computer
The test set needs multiplication	The test set does not need multiplication
It is difficult to anticipate leaking	It is relatively easy to anticipate leaking
It is difficult to secure and to send	It is easy to secure and to send
Time allocation and the number of questions have been decided before, so banks of questions are not needed, and the scoring process requires some time	Time allocation and the number of questions have been decided before, so banks of questions are not needed, and the scores can be obtained immediately

Besides all the strength that have mentioned above, CBT also has some weaknesses, i.e. it costs bigger in the early process of the development and the maintenance and technological facilities have always to be available in every situation, for example, the availability of the internet when an online test is conducted and the test taker has to be able to operate the computer.

In addition, there are some conditions that need to be fulfilled so that CBT can be conducted well, i.e. the availability of bank of questions whose level of difficulty is various completed with the test question's characteristics; the availability of computers in sufficient numbers; the availability of internet connection with enough bandwidth; and the availability of supportive application programs.

### **3 Material and Methodology**

This research is a development research, it aims to develop a product in the form of mathematical formula that is simultaneous model between response accuracy and response time for power test. This model is developed by implementing joint distribution concept, that is a multiplication of conditional distribution model of response accuracy with response time, which is the two-parameter logistic model with marginal distribution model of the response time. At this point, response time is assumed as a random variable, since the response time is not only controlled by the test administration officer or computer, but also other factors that influence the speed of a test taker in answering one test question. The chosen marginal distribution is lognormal distribution because it has positive value and it is in the form of positive skewed in accordance with the response time characteristics and it is easy to present compared to weibull and gamma distribution.

The developed mathematical model is arranged by the parameters that influence the probability of a test taker in answering the test question correctly, such as test taker's parameter and test question's parameter. Parameters related to test questions are the discriminating power of the question ( $a_j$ ), delay due to test questions ( $d_i$ ) and the level of difficulty of the test questions ( $b_j$ ), while parameters related to the test taker are the test taker's ability ( $\Theta_i$ ) and response time ( $t_{ij}$ ). The implementation of this response time should be more realistic compared to that of IRT model that does not involve response time parameter, as in reality, a test always has time constraint even though not all test providers record this information.

The developed product should be suitable to be implemented in power test, so suitability test using empirical data needs to be conducted. Data used in analyzing the simultaneous model between response accuracy and response time is the score of correct answers (1) - wrong answers (0) and the test taker's response time taking quiz of statistics, descriptive statistics materials using CBT provided by the lecturers in Institut Agama Islam Negeri Ponorogo that consist of 10 multiple choice questions. This quiz is done by 572 students of Institut Agama Islam Negeri Ponorogo from various majors and study programs who take Statistics in the even semester in 2017.

The simultaneous model is developed using joint distribution concept. The following are the procedures:

- Formulating conditional distribution model of response accuracy with response time (two-parameter logistic model) based on the related theoretical review
- Formulating response time model based on the appropriate marginal distribution characteristics (modified lognormal distribution)
- Multiplying conditional distribution of response accuracy by response time with marginal distribution to obtain joint model.

After the simultaneous model is developed, parameter estimation is conducted. Parameter estimation is done using Bayesian Markov Chain Monte Carlo (MCMC). This method is considered appropriate to simultaneous model with complex parameters as, according to Natesan [9], Bayesian method is able to estimate model parameters that are very complex, violate IRT basic assumptions, and have small size sample. The steps done to estimate the parameters are:

- Determining likelihood distribution based on the data
- Determining prior distribution that is suitable with every parameter in the model
- Determining posterior distribution
- Implementing Gibbs Sampler algorithm to rise the parameter value based on the obtained posterior distribution. The implementation of the algorithm is done with the assistance of WinBUGS 1.4 program.

After the value of all parameters is estimated, the suitability test on the developed simultaneous model using empirical data, like what has been explained previously, is conducted. The method used is Posterior Predictive Model Checking (PPMC), by comparing the difference between the data obtained from the observation and the data from the prediction model result (discrepancy) based on the correct answer score using a graph. The smaller the difference is, the more suitable the model to model the data.

The last procedure is analyzing the estimation result of model parameter. The purpose of this analysis is to identify the characters of the test takers and test questions.

## **4 Results and Discussion**

In this research, response time is assumed as a random variable. It means time response time is not only controlled by the test administration officer or computer, but also other factors that influence the speed of a test taker in answering the test question, so the required time has a certain distribution based on their characteristics. According to Lindsey [14], there are some characteristics worth considering in determining response time distribution:

- It has to have positive value
- Positive skewed (the amount of probability of short response time is bigger than the probability of long response time). There are several distributions that are considered appropriate to the response time characteristics, i.e lognormal distribution, weibull and gamma. In this research, lognormal distribution is chosen as a marginal distribution because among those three distributions, lognormal distribution is the

easiest distribution to interpret. There are some experts who develop the simultaneous model using marginal lognormal distribution, like Thissen [29], Ingrisone [24], Hidayah and Kumaidi [17]. However, the lognormal distribution structure developed by those experts is different from each other (the mean definition), as the structure of lognormal distribution used in this research, which is a modification of lognormal distribution developed by Thissen [29] dan Ingrisone [24].

Lognormal distribution shows the relation between response time variable and other influential variables. Therefore, response time is a dependent variable and other influential variables are independent variables. Response time is influenced by many factors. Halkitis [30] in their research found that there is a positive relationship between response time and the questions' level of difficulty, the question's discriminating power, and the question's length (time required to read the question). Then, Zenisky and Baldwin [31] study the relationship of response time with the question's level of difficulty, questions' level of complexity, question's content and the difference of question group. Based on those factors, mean of lognormal distribution can be redefined to improve the existing model so that it can be more realistically implemented.

By modifying the lognormal distribution developed by Thissen [29] and Ingrisone [24] and considering factors that influence the response time, the mean value of natural log response time can be formulated as  $\mu_{\ln t_{ij}} = v + d_i + g \cdot z_{ij}$ . Natural log response time has normal distribution, or in other words, response time distribution has lognormal distribution, so  $\ln t_{ij} \approx N\left(v + d_i + g \cdot z_{ij}, \sigma^2\right)$ , or can be formulated as:

$$f(t; \theta, a, b, v, d, g) = \frac{1}{t_{ij} \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma}\right]^2\right) \tag{12}$$

where  $z_{ij} = a_j \theta_i + b_j$

In the mentioned formula above, mean of all response times is symbolized as  $v$ . The delay due to test questions is symbolized as  $d_i$ . The bigger the delay parameter due to the test questions, the longer the response time needed. This parameter is not related to the delay because of the question's level of difficulty, but the delay because of the number of words in the test question or the delay because of the test question is in the form of picture or graph like that in Statistics subject. Regression coefficient shows the relationship between ability ( $\theta_i$ ) and the level of difficulty ( $b_j$ ) and response time ( $t_{ij}$ ) is symbolized as  $g$  and variation of  $t_{ij}$  is symbolized as  $\sigma^2$ .

The formula of two-parameter logistic model integrated with response time that will be developed in this research is the modified results of Hidayah and Kumaidi [17]. The following is the model.

$$p(X = 1 | \theta_j, t_{ij}, a_j, b_j) = \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \tag{13}$$

Based on the model above, the probability of the test taker answering the test questions correctly is influenced by the test taker’s ability that is symbolized as  $\theta_i$ , the test question’s discriminating power that is symbolized as  $a_j$ , the question’s level of difficulty that is symbolized as  $b_j$ , the delay due to the test question’s factors that is symbolized as  $d_j$  and response time that is symbolized as  $t_{ij}$ .

To obtain the simultaneous model between response accuracy and response time, joint distribution concept is implemented. Joint distribution is formulated in the following formula:

$$f(x_{ij}, t; \theta, a, b, v, d, g) = f(x_{ij} | t; \theta, a, b, d) \cdot f(t_{ij} | \theta, a, b, v, d, g) \tag{14}$$

Conditional distribution of response accuracy and response time is formulated as follows:

$$f(x_{ij} | t; \theta, a, b, d) = \left[ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{x_{ij}} \left[ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1 - x_{ij}} \tag{15}$$

Marginal distribution of response time is as follows:

$$f(t_{ij} | \theta, a, b, v, d, g) = \frac{1}{t_{ij} \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[ \frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma} \right]^2\right) \tag{16}$$

Joint distribution is as follows:

$$f(x_{ij}, t; \theta, a, b, v, d, g) = \left\{ \left[ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{x_{ij}} \left[ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1 - x_{ij}} \right\} \cdot \left\{ \frac{1}{t_{ij} \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[ \frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma} \right]^2\right) \right\} \tag{17}$$

Parameter estimation method used in this research is Bayesian. In this method, all parameters in the model is random variables, so all of them is stated in the form of distribution. Parameter estimation is done by determining prior distribution that suits every parameter in the model and likelihood distribution that suits the previous data to obtain posterior distribution.

Prior distribution is chosen based on the parameter domain that is based on the theory [32, 33]. The test taker's ability is in the range of -3 and 3, so the prior distribution used is normal,  $\theta_i \sim N(0,1)$ . The question's discriminating power is in the range of -3 and 3, so the prior distribution used is normal,  $a_j \sim N(0,1)$ . The question's level of difficulty is in the range of -3 and 3, so the prior distribution used is normal,  $b_j \sim N(0,1)$ . The delay due to question's factors always has positive value, so the prior distribution used is truncated normal,  $d_j \sim N(0,1); d_j > 0$ . The mean of response time is related to the length of time required to answer the question, so the prior distribution used is gamma distribution,  $v \sim \Gamma(3,2)$ . Regression coefficient that reflects the relationship between the test taker's ability and the question's level of difficulty with the response time has prior uniform distribution,  $g \sim U(0.3,0.7)$ , because the value of each random variable  $g$  has the same probability. The identified variation also has prior uniform distribution,  $\sigma^2 \sim U(0,1)$ .

Likelihood distribution is determined based on data, that is the test taker's answers in the form of dicotomy in which the score for correct answer is 1 and the score for wrong answer is 0, so the appropriate likelihood distribution is bernoulli distribution.

The likelihood distribution is as follows.

$$L(x_{ij}, t_{ij} | \theta, a, b, v, d, g) = \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left\{ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{1-x_{ij}} \left\{ \frac{1}{t_{ij} \sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[ \frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))^2}{\sigma} \right]^2\right) \right\} \quad (18)$$

Posterior distribution is the proportion of the multiplication between likelihood distribution and prior distribution.

Posterior distribution for parameter  $\theta_i$  can be formulated as follows.

$$f(\theta | x_{ij}, t, a, b, v, d, g) = \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left\{ \frac{\exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j \left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{1-x_{ij}} \right\}$$

$$\left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) \right] \quad (19)$$

Posterior distribution for parameter  $a_j$  can be formulated as follows.

$$f(a|x_{ij}, t, \theta, b, v, d, g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{a_j^2}{2}\right) \right] \quad (20)$$

Posterior distribution for parameter  $b_j$  can be formulated as follows.

$$f(b|x_{ij}, t, \theta, a, v, d, g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{b_j^2}{2}\right) \right] \quad (21)$$

Posterior distribution for parameter  $d_j$  can be formulated as follows.

$$f(d|x_{ij}, t, \theta, a, b, v, g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (v + d_i + g \cdot z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{d_j^2}{2}\right) \right] \quad (22)$$

Posterior distribution for parameter  $\nu$  can be formulated as follows.

$$f(\nu|x_{ij},t,\theta,a,b,d,g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (\nu + d_i + g.z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\nu^2}{2}\right) \right] \quad (23)$$

Posterior distribution for parameter  $g$  can be formulated as follows.

$$f(g|x_{ij},t,\theta,a,b,d,g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (\nu + d_i + g.z_{ij}))}{\sigma}\right]^2\right) \right\} \left[ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{g^2}{2}\right) \right] \quad (24)$$

Posterior distribution for parameter  $\sigma$  can be formulated as follows.

$$f(\sigma|x_{ij},t,\theta,a,b,\nu,d,g) = \left[ \prod_{I=1}^N \prod_{j=1}^I \left\{ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right\}^{x_{ij}} \left[ \frac{\exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - \frac{d_j}{t_{ij}} - b_j\right)\right)} \right]^{1-x_{ij}} \right] \left\{ \frac{1}{t_{ij}\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\ln t_{ij} - (\nu + d_i + g.z_{ij}))}{\sigma}\right]^2\right) \right\} \quad (25)$$

Posterior distributions above are models of complex parameters, so the numerical iteration uses MCMC. To rise the parameter value based on the mentioned posterior distributions above, Gibbs Sampling algorithm is used. The implementation of that algorithm is done with the assistance of WinBUGS 1.4 program. The results of the

parameter estimation can be used if the iteration has reached convergent condition. This condition can be seen in ACF, history and density.

After developing simultaneous model between response accuracy and response time and all parameters have been estimated with the assistance of WinBUGS 1.4 program, the suitability test using empirical data needs to be conducted. This test aims to identify whether the developed simultaneous model is suitable for power test. As explained previously, the method used in the suitability test is Posterior Predictive Model Checking (PPMC) with graphs. In PPMC, the most important thing to do is the deciding the discrepancy size (the difference between data obtained from the observation and data from the model prediction result). In this research, the chosen discrepancy size is the distribution of total score. According to Chon, Nozawa and Zhang [34], the most effective discrepancy size to identify the model suitability is the total score distribution. The model is in accordance with the data if the line of total score distribution of the data prediction on the graph is almost close to the line of total score distribution of the observation data. The following is the comparison graph of the total score distribution (Fig. 1).

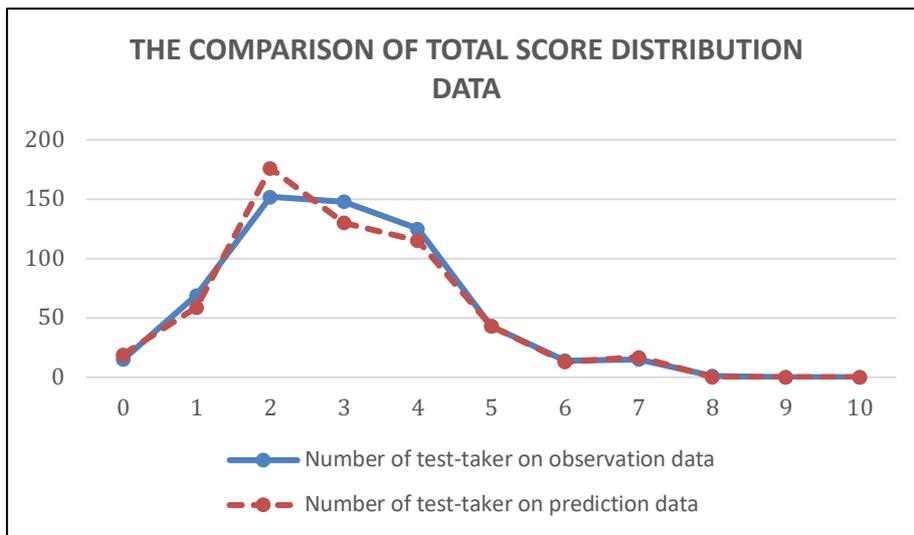


Fig. 1. The Comparison of Total Score Distribution Data

From the graph, the lines of the total score distribution of observation data and prediction data are very close. It shows that the simultaneous model is appropriate to be used in power test.

Next, the results of estimation parameter of simultaneous model using WinBUGS Codes are as follow (Table 2). The estimation results of discriminating power parameter (a) in the Table 2 are in the range of 0.6438 and 0.7201. It shows that the question’s ability to differentiate the test takers’ high and low ability is fair, as the average score is close to 0.

**Table 2.** The Estimation Results of Parameter a

Node	Mean	SD	MC Error	Median
a[1]	0.6843	0.0355	0.0025	0.6830
a[2]	0.7201	0.0386	0.0025	0.7183
a[3]	0.6662	0.0352	0.0023	0.6650
a[4]	0.6848	0.0363	0.0024	0.6828
a[5]	0.6979	0.0361	0.0023	0.6962
a[6]	0.6697	0.0344	0.0023	0.6678
a[7]	0.6802	0.0352	0.0025	0.6784
a[8]	0.6438	0.0365	0.0023	0.6423
a[9]	0.6808	0.0372	0.0026	0.6780
a[10]	0.6840	0.0380	0.0025	0.6824

**Table 3.** The Estimation Results of Parameter b

Node	Mean	SD	MC Error	Median
b[1]	1.403	0.1040	0.0062	1.403
b[2]	1.207	0.1241	0.0084	1.208
b[3]	1.404	0.1309	0.0090	1.406
b[4]	1.379	0.1160	0.0076	1.381
b[5]	1.363	0.1098	0.0064	1.362
b[6]	1.449	0.1208	0.0081	1.453
b[7]	1.445	0.1116	0.0075	1.442
b[8]	1.461	0.1433	0.0099	1.465
b[9]	1.439	0.1286	0.0092	1.438
b[10]	1.320	0.1397	0.0098	1.323

Meanwhile, the estimation results of difficulty level parameter (b) based on the Table 3 show that the value is in the range of 1.207 and 1.461. It shows that the given test questions are categorized as difficult, because the average score is close to 2.

The estimation results of delay due to question factors parameter (d) are in the range of 0.0521 and 0.1238. This estimation value shows that the test taker’s delay because of the test questions factors is categorized as fair as the score is close to 0. Delay parameter due to this factor is identified on the questions that consist of picture or graph as those in Statistics subject.

**Table 4.** The Estimation Results of Parameter d

Node	Mean	SD	MC Error	Median
d[1]	0.0521	0.0432	0.0017	0.0415
d[2]	0.1224	0.0833	0.0057	0.1105
d[3]	0.0895	0.0696	0.0047	0.073
d[4]	0.0795	0.0600	0.0033	0.0683
d[5]	0.0651	0.0509	0.0024	0.0536
d[6]	0.0716	0.0568	0.0028	0.0583
d[7]	0.0531	0.0459	0.0022	0.0407
d[8]	0.0979	0.0709	0.0046	0.0848
d[9]	0.0774	0.0613	0.0036	0.0639
d[10]	0.1238	0.0795	0.0049	0.1122

If theta value is arranged from the lowest to the highest, the graph will look like the following (Fig 2). Based on the graph on Fig 2, it can be concluded that the estimation (theta) of students' ability in Institut Agama Islam Negeri Ponorogo to Statistics subject, Statistics descriptive materials in the scale of interbal (-3.3) ranges between -2.53 and 1.2, so it is categorized as low until high. However, to obtain more precise result, it requires linier transformation, for example on the scale of 1 to 10 or on the scale of 1 to 100.

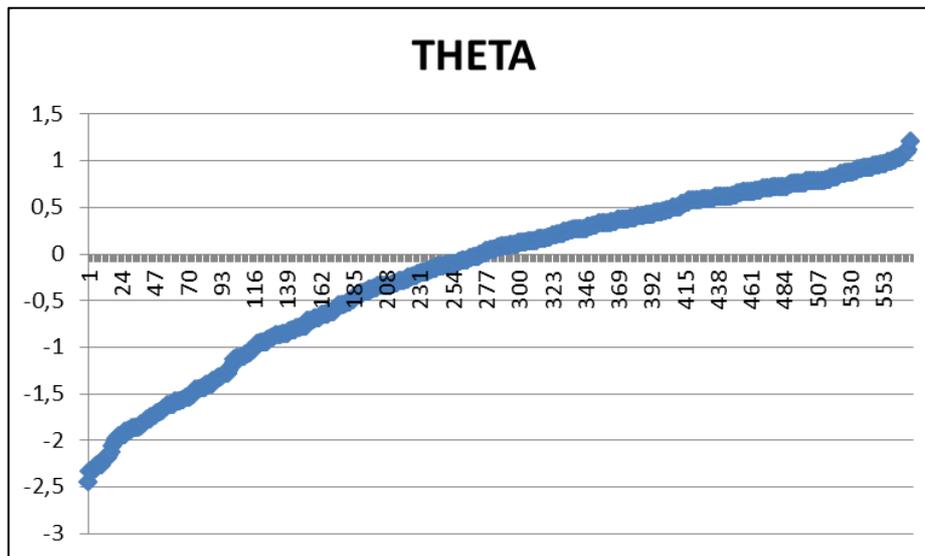


Fig. 2. The Estimation Results of Theta Parameter

## 5 Conclusion

Based on the results and discussion, it can be concluded that simultaneous model developed by multiplying two-parameter logistic model integrated with response time by lognormal response time model is the most suitable model for CBT power test. However, the conclusion of this suitability test is limited to the use of empirical data. For the next research, it is suggested to use simulation data with various scenarios so that the data can be compared on how it is suitable for any conditions. Meanwhile, the parameter estimation results using Bayesian method on empirical data also lead to some conclusions. It can be identified that the estimation results of discriminating power parameter show that the question's ability to differentiate the test taker's ability level is categorized as fair, the estimation results of difficulty level parameter show that the given test questions are categorized as difficult, the estimation results of delay parameter due to question factor are categorized as fair, and it is considered as reasonable because the questions are related to pictures and graphs. Finally, the ability estimation (theta) of the test takers is categorized as low to high.

## 6 Acknowledgement

The authors thank to Institut Agama Islam Negeri Ponorogo for the empirical data.

## 7 References

- [1] Lee, Y.-H. and Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*, 53(3): 359.
- [2] Verbić, S. (2010). Establishing the rhythm of responding to test questions.
- [3] Hadadi, A. and Luecht, R. M. (1998). Some methods for detecting and understanding test speededness on timed multiple-choice tests. *Academic Medicine-Philadelphia-*, 73 S47-S50. <https://doi.org/10.1097/00001888-199810000-00042>
- [4] Dodonova, Y. A. and Dodonov, Y. S. (2013). Faster on easy items, more accurate on difficult ones: Cognitive ability and performance on a task of varying difficulty. *Intelligence*, 41(1): 1-10. <https://doi.org/10.1016/j.intell.2012.10.003>
- [5] Schnipke, D. L. and Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. *Computer-based testing: Building the foundation for future assessments*, 237-266.
- [6] Hornke, L. F. (2000). Item response times in computerized adaptive testing. *Psicológica*, 21(1):
- [7] Abdelfattah, F. A., "Response latency effects on classical and item response theory parameters using different scoring procedures," Ohio University, 2007.
- [8] Entink, R. K., "Statistical Models for Response and Response Time," Ph.D., University of Twente, Enchede, 2009.
- [9] Natesan, P., "A review of bayesian item response modeling: Theory and applications," P. Baldwin, Ed., ed: SAGE Publications Sage CA: Los Angeles, CA, 2011.
- [10] Levy, R. (2009). The rise of markov chain monte carlo estimation for psychometric modeling. *Journal of Probability and Statistics*, 2009. <https://doi.org/10.1155/2009/537139>
- [11] Gulliksen, H. (2013). *Theory of mental tests*. Routledge.
- [12] van der Linden, W. J. and Hambleton, R. K. (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
- [13] Rouder, J. N., Province, J. M., Morey, R. D., Gomez, P., and Heathcote, A. (2015). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, 80(2): 491-513. <https://doi.org/10.1007/s11336-013-9396-3>
- [14] Lindsey, J. K. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press.
- [15] Ranger, J. and Kuhn, J.-T. (2012). A flexible latent trait model for response times in tests. *Psychometrika*, 77(1): 31-47. <https://doi.org/10.1007/s11336-011-9231-7>
- [16] Ingrisone II, J. N., "Modeling the joint distribution of response accuracy and response time," The Florida State University, 2008.
- [17] Hidayah, N. and Kumaidi, K. (2016). Analisis model simultan model logistik satu parameter dengan waktu respon berdasarkan data simulasi. *Jurnal Penelitian dan Evaluasi Pendidikan*, 20(2): 208-220. <https://doi.org/10.21831/pep.v20i2.8068>
- [18] Liu, Y. and Zhao, X. (2017). Design Flow of English Learning System Based on Item Response Theory. *International Journal of Emerging Technologies in Learning (iJET)*, 12(12): 91-102. <https://doi.org/10.3991/ijet.v12i12.7958>

- [19] Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). Fundamentals of item response theory. Sage Publication Inc.
- [20] Roskam, E. E., "Models for speed and time-limit tests," in Handbook of modern item response theory, ed: Springer, 1997, pp. 187-208. [https://doi.org/10.1007/978-1-4757-2691-6\\_11](https://doi.org/10.1007/978-1-4757-2691-6_11)
- [21] Verhelst, N. D., Verstralen, H. H., and Jansen, M., "A logistic model for time-limit tests," in Handbook of modern item response theory, ed: Springer, 1997, pp. 169-185. [https://doi.org/10.1007/978-1-4757-2691-6\\_10](https://doi.org/10.1007/978-1-4757-2691-6_10)
- [22] Wang, T. and Hanson, B. A. (2005). Development and calibration of an item response model that incorporates response time. Applied Psychological Measurement, 29(5): 323-339. <https://doi.org/10.1177/0146621605275984>
- [23] Wang, T. (2006). Model for the joint distribution of item response and response time using a one-parameter weibull distribution. Iowa City: IA: Center for Advanced Studies in Measurement and Assessment,
- [24] Ingrisone, S. J., "An extended item response theory model incorporating item response time," The Florida State University, 2008.
- [25] Zainul, A. and Nasution, N. (2001). Penilaian hasil belajar. PAU untuk peningkatan dan Pengembangan Aktivitas Instruksional Ditjen Dikti, DepDiknas, Jakarta,
- [26] Sudjana, N. (2009). Penilaian Hasil Proses Belajar Mengajar.
- [27] Khoshsim, H. and Toroujeni, S. M. H. (2017). Computer-Based Testing: Score Equivalence and Testing Administration Mode Preference in a Comparative Evaluation Study. International Journal of Emerging Technologies in Learning (iJET), 12(10): 35-55. <https://doi.org/10.3991/ijet.v12i10.6875>
- [28] Georgiadou, E., Triantafyllou, E., and Economides, A. A. (2006). Evaluation parameters for computer-adaptive testing. British Journal of Educational Technology, 37(2): 261-278. <https://doi.org/10.1111/j.1467-8535.2005.00525.x>
- [29] Thissen, D., "Latent trait scoring of timed ability tests," presented at the Computerized Adaptive Testing Conference, Minneapolis, 1979.
- [30] Halkitis, P. N., "Estimating testing time: The effects of item characteristics on response latency," 1996.
- [31] Zenisky, A. L. and Baldwin, P. (2006). Using item response time data in test development and validation: Research with beginning computer users. Center for educational assessment report No, 593
- [32] Nguyen, L. and Do, P., "Combination of Bayesian network and overlay model in user modeling," in International Conference on Computational Science, 2009, pp. 5-14.
- [33] Spiegel, R. (2007). Bayesian statistics as an alternative to gradient descent in sequence learning. International Journal of Emerging Technologies in Learning (iJET), 2(3):
- [34] Chon, K. H., Nozawa, Y., and Zhang, S. (2006). Posterior Predictive Checking of Unidimensional Item Response Theory Models.

## 8 Authors

**Andhita Dessy Wulansari**, female, 35 years old, is a lecturer of Institut Agama Islam Negeri Ponorogo. She received her Doctor of Educational Research and Evaluation at Universitas Negeri Yogyakarta in 2015. She has published several research papers in the area of Statistics and Educational Measurement. She has also authored books, Parametric Statistics and Research Methods in Education. [andhita@iainponorogo.ac.id](mailto:andhita@iainponorogo.ac.id)

**Kumaidi**, male, 67 years old, is a professor at Department of Psychology, Universitas Muhammadiyah Surakarta.

**Samsul Hadi**, male, 58 years old, is a lecturer at Department of Electronics Engineering Education, Universitas Negeri Yogyakarta.

Article submitted 2019-03-31. Resubmitted 2019-06-01. Final acceptance 2019-06-02. Final version published as submitted by the authors.