

Full Length Research Paper

## Item response theory: A basic concept

Jumailiyah Mahmud

Institute of Teaching and Educational Sciences of Mataram, Indonesia.

Accepted 17 January, 2017; Accepted 22 February, 2017

With the development in computing technology, item response theory (IRT) develops rapidly, and has become a user friendly application in psychometrics world. Limitation in classical theory is one aspect that encourages the use of IRT. In this study, the basic concept of IRT will be discussed. In addition, it will briefly review the ability parameter estimation, particularly maximum likelihood estimation (MLE) and expected a posteriori (EAP). This review aims to describe the fundamental understanding of IRT, MLE and EAP which likely facilitates evaluators in the psychometrics to recognize the characteristics of test participants.

**Key words:** Expected A Posteriori, Item Response Theory, Maximum Likelihood Estimation

### INTRODUCTION

Over the last decade, item response theory (IRT) has increasingly been popular. As noted by Steinberg and Thissen (2013), many studies have been conducted to enrich literatures in the field of psychometrics.

It is common to note that IRT is a pivotal methodology which has been globally used in many assessment programs. IRT is commonly applied in educational and psychological testing, and recently it is beneficial to assess health outcomes (Cai et al., 2016).

In educational context, IRT is developed to address the limitation in classic measurement theory, particularly its shortcoming that is dependent between test participant group and items in nature. Such dependent characteristics mean the outcome of the measurement depends on the participant group completing the test. If the test is given to participant group with high ability, the difficulty level of the question item appears to be low. On the contrary, if the test is given to participant group with low ability, the difficulty level of the question item turns

out to be high (Hambleton et al., 1991b).

The estimation of parameters is a central matter in the item response theory, though it is said that the item response theory is successful due to the success of implementing the parameter estimation (Swaminathan, 1983). Matter that strongly needs attention in parameter estimation is large number of empirical data despite its dependency on the model of parameter logistic in use. Based on the aforementioned outline, the writer in this review will describe basic concept of IRT, dichotomous logistic model and the type of ability parameter estimations, particularly that of maximum likelihood and expected a posteriori.

### Item response theory (IRT)

The term of IRT in the literature can be found as latent distribution theory, *item characteristic curve* (ICC) and

E-mail: jumailiyah@gmail.com. Tel: +62 370 632082.

Authors agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

*item characteristic function* (ICF). This item characteristic curve is presented in an item characteristic relation curve with participant characteristics which is shown on the abscissa while the ordinate shows the probability of the item answer.

The test participant characteristics and item characteristics are related by model in the form of function or graphical curve (Naga, 1992b). Each question item is represented by an ICC showing the relation between correct answer probability and the test participant ability.

In classic theory, the item characteristics will depend on the ability level of the test participants, if the item is completed by participant with high ability, the item shows low difficulty level, in contrast, for participant with low or medium ability, the item will show high difficulty level. On the other hand, item response theory predicts the participants' ability from their ability in answering the test items correctly, the higher their ability, the higher the probability of correct answer they provide. Likewise, the higher the item difficulty level, the higher the test participants' ability to answer the item correctly. Despite a claim stating that modern theory cannot substitute classic theory (Zanon et al., 2016), based on the aforementioned description, the basic concept of item response theory is considered a strong theory compared to that of classic theory.

Moreover, recent technology development has made IRT implementation far easier. Yet, the theory requires general assumptions or conditions of item response theory to satisfy by the items and the test participants including:

1. Unidimensional
2. Local independency, and parameter invariance.

Unidimensional specifically means an exam measure only one characteristic of the participants (Crocker and Algina, 1986). Firstly, unidimensional means the exam only measure one character or one ability of the test participants. For instance, one set measures ability in calculation and does not disclose the test participant ability in understanding or mastering language. Statistically, unidimensional can be calculated with factor analysis indicated by one dominant factor.

Secondly, local independency means that the influence of participant ability and test item are considered constant in which the participants' response to question item have no relation statistically. "This assumption will be satisfied when the participants' answer to one item does not influence the answer to another item. The participants' answer to several test items is expected to have no correlation" (Hambleton et al., 1991c). The implication of this assumption results in items analyzable item per item, and likewise the participants are analyzed per individual.

Thirdly, parameter invariance that is, "the function of

item characteristics is constant or remains unchanged albeit the participant group answering the items changes. In the same group, their characteristics will remain unchanged despite the items they answer change" (Naga, 1992b). The invariance is reviewed from the point of item characteristics and the participant characteristics, difficulty level and distinguishing capacity of the items will remain notwithstanding the question items are answered by high ability group or low ability group. The participant ability will be constant or remain unchanged despite the items they answer change.

The most essential assumptions in item response theory are unidimensional and local independency (Embretson and Reise, 2000a). This opinion was also proposed by Hambleton (1989). One of the most common assumptions is that in any test, only one ability is measured by the items instrument. This assumption is called unidimensional assumption (Azwar, 2004).

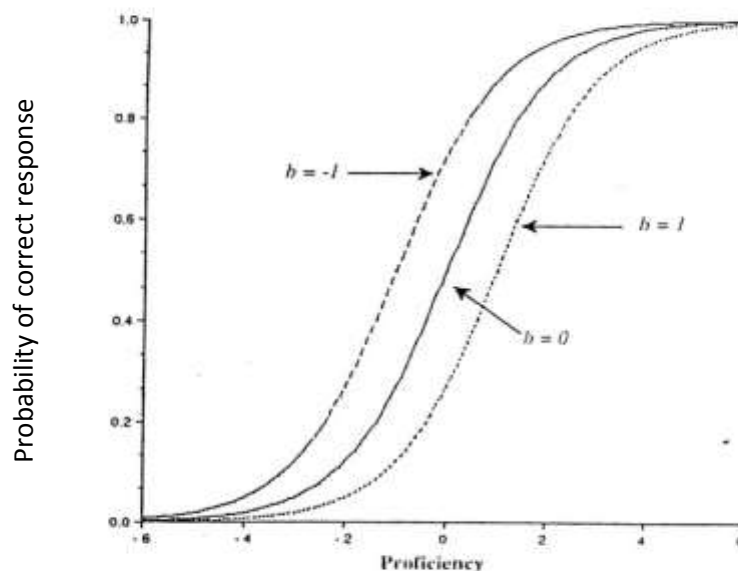
### Dichotomous logistic model

Furthermore, the advantage of item response theory in relation with the analysis of the test result is to present the basis for making prediction, estimation or conclusion on the participants' ability. The process of education measurement starts with scoring the item response of the participant and response pattern matrix is developed, carrying out initial check on the data conformity by choosing the parameter model, estimating the item parameter and the participants ability, and composing the scaling transformation (Hambleton and Swaminathan, 1985)

Some types of data are analyzed with item response analysis model such as dichotomous, polytomous and continuous data. In dichotomous data, response to one item is shown in two categories such as: true-false, yes-no, agree-don't agree. Particularly, the participants' ability test consists of two categories with true or false content. In a test with multiple choice format of five answers option, the categorization of respondents answer will be grouped into two response categories that is, true or false where correct response will score one and incorrect response will score zero (Bejar, 1983)

There are three types of logistic model of item response theory that is, single parameter model, dual parameters model, and triple parameters model. The three models differ in the number of parameter to calculate in describing the item characteristics. The single parameter model calculates only the item difficulty level ( $b_i$ ), while the distinguishing capacity of item ( $a_i$ ) scores one or constant and the guessing parameter scores ( $c_i$ ) zero. The dual parameters model calculates the item difficulty level ( $b_i$ ) and the item distinguishing capacity ( $a_i$ ), while the guessing parameter ( $c_i$ ) scores zero.

Whereas, the triple parameters model calculates the



**Figure 1.** Item curve of single parameter logistic model (Hambleton et al., 1991c).

three parameters that is,  $b_i$ ,  $a_i$  and  $c_i$ .

### Single parameter logistic model

The description of single parameter model in an item characteristics curve,  $b_i$  parameter is a parameter location shown in ability scale, named item difficulty level. The higher the item difficulty level, the more to the right its position in a curve (Lord, 1980b) (Figure 1).

In Figure 1, item difficulty level parameter ( $b_i$ ) is a point in the ability scale to have the opportunity of 50% to answer correctly. There are three items: the third items show the most right item, and therefore they are the items with the highest difficulty level ( $b_i = 1.0$ ), while the first items show the left most items which are the items with the lowest difficulty level ( $b_i = -1.0$ ), and the second items are the items with medium difficulty level that is,  $b = 0$ .

To be able to answer 50% of the first items correctly, participant ability of minimal  $-1.0$  is necessary ( $\theta = -1.0$ ). To be able to answer 50% of the third items correctly, participant ability of minimal  $1$  is necessary.  $0$  ( $\theta = +1.0$ ). The higher  $b_i$  scores, the higher the ability necessary to answer 50% of them correctly. The higher  $b_i$  scores, the more difficult the items or otherwise, the lower  $b_i$  scores, the easier the items.

When the ability value is transformed with mean = 0 and deviation standard = 1, the  $b_i$  value will vary from  $-2.0$  to  $+2.0$ .  $b_i$  value close to  $-2.0$  indicates low difficulty

level of the items while  $b_i$  value close to  $+2.0$  indicates high difficulty level of the items or difficult question item (Hambleton et al., 1991c).

### Dual parameter logistic model

In single parameter model, the form of the curve is similar while the dual parameter logistic model calculates the curves with different curve slopes. The difference in items curve indicates the difference of  $a_i$  value. Items with steep curve indicates high distinguishing capacity or high  $a_i$  value, while gradient item curve indicates low difficulty level or low  $a_i$  value. Subsequently, a chart is presented showing two item characteristics with equal and different distinguishing capacity ( $a_i$ ) and difficulty level ( $b_i$ ) (Embretson and Reise, 2000b) (Figure 2).

Figure 2 shows 4 items with equal and different difficulty level, and distinguishing capacity. The first and fourth item have equal difficulty level that is,  $b_i = 1$ , this can mean that both items requires participants' ability of  $\theta = 1.0$  to be able to answer 50% of the items correctly.

However, both items have different distinguishing capacity that is, the first items with  $=0.5$  while the fourth item is  $a_i = 1$  showing more gradient curve on the first items than that of the fourth's, the third items shows steep curve due to high distinguishing capacity ( $a_i = 3$ ), and the second items on the left most indicate lowest difficulty level ( $b_i = -1$ ) having the same gradient with the fourth items since they have the same distinguishing capacity ( $a_i = 1$ ).

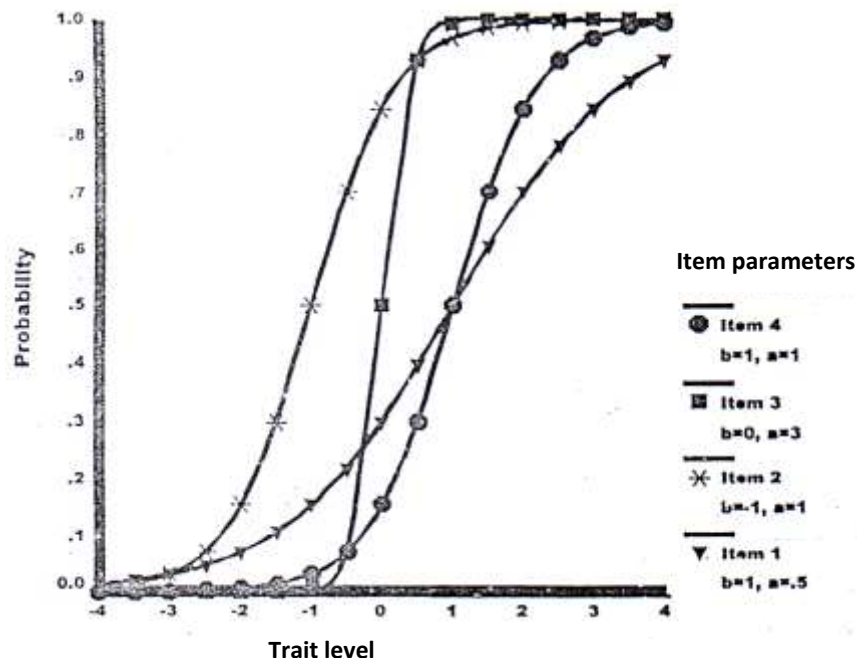


Figure 2. Item Curve of Dual Parameter Logistic Model (Embretson and Reise, 2000b).

Theoretically, parameter  $a_i$  is located at scale  $-\infty$  to  $+\infty$ . However, a difficult case occurs to  $a_i$  negative value, or it is better to remove items with negative distinguishing capacity due to possible error, and this indicates correct answer probability decreases when the ability level increases.  $a_i$  is also unlikely to occur if it is bigger than 2.0. The  $a_i$  value commonly ranges from 0.0 to 2.0. Dual parameter logistic model formula are:

$$P_i(\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}} \quad i=1,2,3 \quad (1)$$

$P_i(\theta_j)$  = the j test participant probability with ability

answer ( $\theta_j$ ) the item i correctly (item  $i = 1, 2, 3... n$ )

$b_i$  = difficulty level parameter of item i

$a_i$  = distinguishing capacity of item i

$n$  = number of item in the test.

$e$  = number with value of 2.718

$D$  = scaling factor made in order the logistic function close to ogive normal function (Embretson and Reise, 2000d).

Furthermore, to find out the items value ( $a_i$  and  $b_i$ ) and the participants ability ( $\theta$ ) in logistic model, we need to

figure out the parameter value in a measurement. The determination of parameter value is known as parameter estimation, item parameter and ability parameter; participant ability value estimation is called scoring and item parameter estimation is called calibration.

### Ability parameter estimation

#### Item response theory ability estimation

Estimation means the process of estimating or predicting. "The estimation contains the finding of value according to parameters of an expression with certain methods" (Makridakis et al., 1999). Estimation is made on regression model and item response theory model, yet they have the following differences:

1. Regression model is commonly applied to variable with linear relation, while parameter logistic model comes with nonlinear relation between question items and participant ability.
2. Independent variable in the regression is a variable that can be observed while in item response theory, the participants ability independent variable ( $\theta$ ) cannot be observed (Hambleton et al., 1991a). Since actual value of item parameter and participant ability are unknown, analysis and estimation

on respondent ability and items parameter are carried out.

At its emergence in 1950s, item response theory did not gain popularity due to the lack of worthy statistical estimation procedure (Birnbaum, 1968). After analysis made by using computer, estimation carried out by Bock and Lieberman (1970) shows that computer-made estimation can be done despite its limitation, that is, analysis is carried out on small number of question items and participants sample. The estimation of parameters is a central matter in the item response theory, even it is said that the item response theory is successful due to the success of implementing and procedure availability that is up the mark of parameter estimation (Chen W-H and Thissen, 1999).

Item parameter estimation is with the assumption that the participants' ability is known or otherwise, and the item parameter is known to estimate the participants' ability. The participants' ability estimation with test instrument where the items have been calibrated; the questions whose item characteristics have calculated will be saved in item bank and will be re-used according to the objectives and information function of the target test.

"Ability estimation procedure can be performed with maximum likelihood (ML), maximum a posteriori (MAP), and expected a posteriori (EAP), while items estimation can be performed with estimation approaches including Maximum Likelihood (ML), Bayesian, logistic regression and heuristic estimation" (Embretson and Reise, 2000c).

Ability estimation with maximum likelihood method is carried out through calculation process with various scoring algorithm. The researcher focuses on two methods that is, MLE and EAP, and provides the details for each method. Estimation method description viewed from theoretical point to facilitate the understanding. Based on the writer's experience, the analysis obtained with MLE and EAP methods is not exactly the same. This may be influenced by factors in data sampling (Mahmud et al., 2016).

Ability estimation and item estimation can be carried out simultaneously that is, an estimation process in which item parameter estimation is carried out with ability estimation. The first step is to estimate the item parameter, and the result will be used to estimate the ability parameter before using it as a value to estimate the item parameter in the subsequent stage.

Henceforth, iteration is carried out in which the value obtained in the current round will be taken as the initial value for the subsequent round; this iteration is carried out until the value difference between one round and another becomes a shade of difference called convergent. The iteration process will stop when it reaches the convergent, and the parameter value obtained on the convergent will become the parameter estimation value to figure out.

### Maximum likelihood estimation (MLE)

Maximum likelihood is a common method for model parameter estimation, sufficiently effective with large sample and valid model application (Longford, 2008). The "likelihood" means probability or possibility, while "maximum" means the highest extent.

Therefore, "maximum likelihood" is the occurrence with the highest possibility. In the literatures on item response theory, "Maximum likelihood is a mode of total likelihood" (Bock, 2003). The highest opportunity will depend on the probability of the correct answers and incorrect answers by the participants, and also on the logistic parameter employed, thus, the determination of maximum ability value is carried out through iteration calculation (Baker, 2001).

Ability estimation with maximum likelihood method is a calculation process which aims to figure out maximum ability ( $\theta$ ) value of each participant with the symbol of  $L_i(\theta)$  and value of  $P_j(\theta)$  obtained from formula 1. Afterwards, the calculation process with formula 2 to formula 5 that can be carried out to calculate ability estimation with maximum likelihood method through iteration process in Bilog MG program is:

$$\log L_i(\theta) = \sum_{j=1}^n \{x_{ij} \log_e P_j(\theta) + (1 - x_{ij}) \log_e [1 - P_j(\theta)]\} \quad (2)$$

$L_i(\theta)$  = Maximum ability value for each participant

$P_j(\theta)$  = Ability probability on each item such as shown in formula 1.

$x_{ij}$  = the number of correct items

$x_{ij}$  = the number of incorrect items

The maximum likelihood value of each participant's ability  $L_i(\theta)$  is derived in the form of logarithm equalized to zero with the following formula:

$$\frac{\partial \log L_i(\theta)}{\partial \theta} = \sum_{j=1}^n \frac{x_{ij} - P_j(\theta)}{P_j(\theta)[1 - P_j(\theta)]} \frac{\partial \log(\theta)}{\partial \theta} = 0 \quad (3)$$

ML estimation,  $\hat{\theta}$ , is calculated with Fisher scoring method named "Fisher information". Theoretically, the method that can be employed in iteration technique varies. Yet, common methods to employ are Newton-Raphson and Fisher. The logic in both approaches does

not have significant differences in which Fisher method employs "Fisher Information" while Newton-Raphson employs "Hessian" as the second "partial derivative" (Brown, 2014). For simplification purpose, the writer employs Fisher method with information function formula for dual parameter model as follows:

$$I(\theta) = \sum_{j=1}^n a_i^2 P_j(\theta) [1 - P_j(\theta)] \tag{4}$$

$I(\theta)$  = information function of respondent's ability  
 $\theta$  = respondent's ability level  
 $a_i$  = distinguishing capacity of item i

Upon obtaining the ability information score of dual logistic parameter model, the iteration can be determined with the following formula:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + I^{-1}(\hat{\theta}) \left( \frac{\partial \log L_i(\hat{\theta})}{\partial \theta} \right) \tag{5}$$

$\hat{\theta}_{t+1}$  = ability estimation score in the current round

$\hat{\theta}_t$  = ability estimation score in the previous round

$I(\hat{\theta})$  = information function of the ability

$\left( \frac{\partial \log L_i(\hat{\theta})}{\partial \theta} \right)$  = Maximum likelihood estimation score of a respondent

The calculation is carried out until the ability score does not change in the last round as of the previous round or convergent. The convergent criterion is 0.05 or 0.01, or even less such as 0.001. With the convergent calculation, ability estimation score ( $\hat{\theta}$ ) is obtained.

In fact, constraint in the logistic model is commonly encountered. Theoretically, the curve in logistic model extends towards 0 or 1 asymptotically. This means that the curve will reach 0 or 1 at infinite point that the estimation method is incapable of estimating the parameter when there are items or participants make all correct answer or all incorrect answer (Naga, 1992a).

**Expected A posteriori (EAP) bayes**

Lord (1980a) described possible use of Bayes estimation since education field usually "give test to the same participant year to year with parallel test or similar test.

Thereby, good description on ability frequency distribution in participants group can be represented".

Birnbaum paradigm is employed to estimate item parameter, and participant parameter individually and simultaneously as implemented in LOGIST program. In Bilog or Bilog MG program, estimation is made in two stages, the first stage is items estimation followed by ability estimation. Data analyzed in maximum likelihood estimation is the data from the participants' response or sample data. Bayes estimation procedure uses prior data and sample data which is used in maximum likelihood approach. During the initial use of Bayes approach, Swaminathan and Gifford (1985) used the hierarchy Bayesian estimation procedure, yet it was complicated to implement due to the lack of computer program available for such purpose. Researchers have adopted more pragmatic approach in which Bayesian approach is considered a tool to improve parameter estimation" (Baker and Kim, 2004a).

Posterior distribution estimation is a combination of prior distribution and sample distribution. The combination based on Bayes rule on conditioned probability often encounters complicated constraint, and this renders difficulty in formulating posterior distribution in statistics while Bayes estimation remains opposed due to the use of prior distribution as researcher's subjective consideration (Baker, 1991).

Despite Bayes estimation is presumed complicated from other review point, item response estimation is employed for its practicality, and complicated calculation in accordance with the new development in computer field that is getting steady and simpler. The estimation does not use integration but is based on discrete distribution in "Mislevy quadrature point".

In fact, Baker, on Bayes estimation stated "the latest advancement in IRT estimation procedure is Bayes estimation implemented for the first time in BILOG program (Mislevy and Bock, 1986), that it will be able to address various issues that comes with simultaneous estimation approach in JMLE method" (Baker, 1991).

EAP ability estimation method estimates respondent ability for response pattern of all correct and all incorrect. This method is part of Bayes approach derived from the average of *posterior* distribution, and does not use any mode. Analysis strategy logic employs Bayes principles in BILOG using Mislevy Histogram, a histogram description showing an area in a curve (Baker and Kim, 2004b). Figure 3 facilitates the understanding in ability estimation with Bayes approach as follows:

1. Determining  $X_k$  ( $k = 1, 2, 3, \dots, q$ ) called *nodes*. Bilog MG default includes 15 nodes.
2. There is density in ordinate area that is, histogram ordinate. Density or weight is usually taken from normal distribution as well as from empirical data.
3. Since BILOG program does not use Hermit-Gauss that

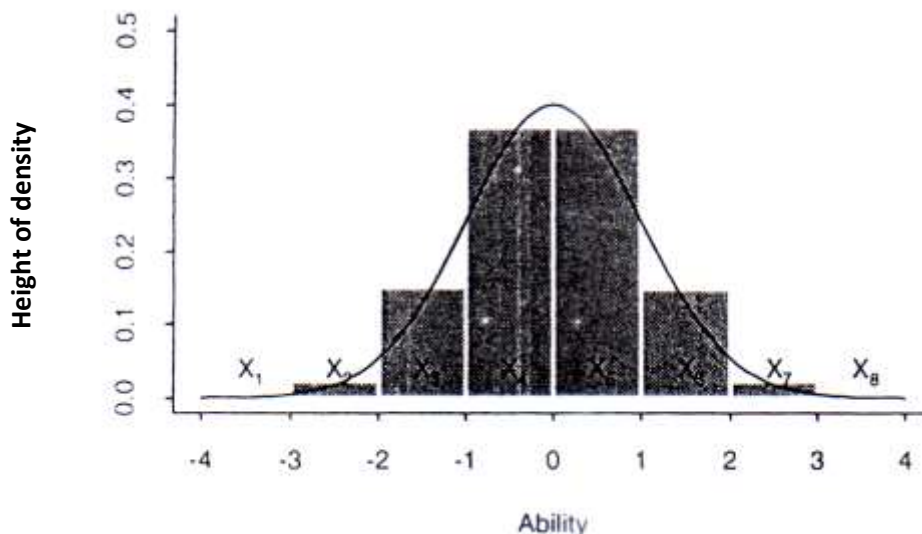


Figure 3. Area in a curve (Baker and Kim, 2004b).

applies integration, but "Mislevy histogram" instead while the it is assumed as normal distribution, then  $X_k$  value can be figured out,  $A(X_k)$  weight shows the gap between  $X_k$  and other  $X_{k+1}$ . If  $X_k$  shows the same gap than  $A(X_k)$  value can be figured out by: one divided into many nodes, otherwise if  $X_k$  does not have the same gap then  $A(X_k)$  is  $X_k$  where  $X_{k+1}$  is called weight.

4. Calculating  $L(X_k)$  value  $P_j(\theta)$  shows ability probability in an item like that obtained from formula 1, while the formula notation is based on Mislevy Histogram using  $P(X_k)$ . Therefore,  $P_j(\theta) = P(X_k)$  and  $X_k = \theta_k$ , while  $L$  is likelihood function of participant ability with formula in the form of multiplication such that that is indicated in formula 6.

$$L(X_k) = \prod_{i=1}^n P_i(X_k)^{u_{ij}} Q_i(X_k)^{1-u_{ij}} \quad (6)$$

$L(X_k)$  = maximum ability value for each participant

- $X_k$  =  $\theta_k$  ability level
- $P_i$  = correct answer probability
- $Q_i$  = incorrect answer probability.
- $u_{ij}$  = the number of correct answer
- $1 - u_{ij}$  = the number of incorrect answer

5. Calculating ability estimation value

$$6. E(\theta_j | U_j, \xi) = \bar{\theta}_j = \frac{\sum_{k=1}^q X_k L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \quad (7)$$

$E(\theta_j | U_j, \xi) = \bar{\theta}_j$  is average ability level, provide that participants' response in dichotomous 0 or 1 scoring is known

$E(\theta_j)$  = expected ability value

$X_k = \theta_k$  = ability level

$A(X_k)$  = weight, indicating gap between  $X_k$  and  $X_{k+1}$

$\xi$  = item parameter value

$q$  = the number of node (quadrature point), the number of group by ability level.

EAP method will be able to analyze or calculate the participant ability notwithstanding that they make all correct answer or all incorrect answer; the calculation process is carried out without iteration but based on average answer score for each participant on answering a number of items.

Relevant to the aforementioned discussion, the working principle of Bayes method starts from posterior data as combination of sample data and prior data or initial data. In education field, prior data can be obtained from data before the study's data collection. In the implementation of item parameter or ability parameter estimation, prior data can be made as artificial data by Bilog MG program using formula 8 and 9 (Baker and Kim, 2004b).

$$\bar{f}_{ik} = \sum_{j=1}^N \left[ \frac{L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \right] \quad (8)$$

$\bar{f}_{ik}$  = artificial examinee for each participant ability ( $X_k$ )

$X_k = \theta_k$  = ability level

$A(X_k)$  = weight, indicating gap between  $X_k$  and  $X_{k+1}$

$q$  = the number of node (quadrature point), the number of group by ability level.

$$\bar{r}_{ik} = \sum_{j=1}^N \left[ \frac{u_{ij} L(X_k) A(X_k)}{\sum_{k=1}^q L(X_k) A(X_k)} \right] \quad (9)$$

$\bar{r}_{ik}$  = artificial data of correct answer in item  $i$  of ability level ( $X_k$ )

$X_k = \theta_k$  = ability level

$A(X_k)$  = weight, indicating gap between  $X_k$  and  $X_{k+1}$

$u_{ij}$  = the number of correct answer

$q$  = the number of node (quadrature point), the number of group by ability level.

The aforementioned discussion includes two ability estimation methods from maximum likelihood approach that is, maximum likelihood estimation method and from Bayes approach that is, expected a posteriori ability estimation method. They differ in:

1. Calculation procedure, formulas used in the calculation. Maximum likelihood method calculates through iteration process while expected a posteriori ability estimation method calculates through average answer of each participant for each ability level.
2. For participants with all correct answer or all incorrect answer in maximum likelihood estimation method, the implementation of joint maximum likelihood estimation (JLME) will not bring calculation result, yet, in the calculation via Bilog MG program, the ability estimation result will be displayed.
3. Maximum likelihood estimation calculation data is based on sample data, while expected a posteriori method uses prior data generated by Bilog MG program using formula 8 and 9 as described earlier.

In this review, it is understood that ability estimate with MLE method implements formula 2, 3, 4, and 5, whereas

the ability estimate with EAP method uses formula 6, 7, 8, and 9. In actual context, formula 6 is quite similar to formula 1, but they have different term and symbol as further implemented with Mislevy Histogram. Formula 8 and 9 are used to create prior artificial data by Bilog MG.

### CONCLUSION

The implementation of IRT as a rigid theory requires requirements test and assumption that the implementation analysis can be further carried out. In general, the assumption does not require any test, however, when assumption test is performed then it can be considered a requirement test even beyond the three assumption tests that is, fitness model test to figure out whether the empiric data (items) is suitable for IRT analysis  $\chi^2$ .

Item characteristics including difficulty level (b), distinguishing capacity (a) and correct answer (c); item characteristics calculated in the analysis influence the mathematics model employed and logistic parameter model. Difficulty level only, indicates its single parameter logistic model; difficulty level and distinguishing capacity indicate dual parameter logistic model; and calculating the three characteristic indicates triple parameter logistic model.

There are three methods that can be used in ability estimation are, MLE of maximum likelihood group, Bayes EAP, and Bayes MAP of Bayes (not discussed) group. Bayes group uses prior data, empirical data and combination of both data, posterior data, in the analysis. Prior data can be generated by BILOG MG program as artificial data using formulas. While theoretically, Maximum likelihood method does not use prior data. This method is said to have no bias elements yet it often fails in ability and items analysis to determine the estimation value; the failure to determine the value on item data all answered correctly and all answered incorrectly.

### CONFLICT OF INTERESTS

The author has not declared any conflict of interests.

### REFERENCES

Azwar S (2004). Dasar-dasar Psikometri. Yogyakarta: Pustaka pelajar.  
 Baker FB (1991). Comparison of minimum logit chi-square and Bayesian item parameter estimation. Br. J. Math. Stat. Psychol. 44(2):299-313. doi: 10.1111/j.2044-8317.1991.tb00963.x.  
 Baker FB (2001). Estimating an Examinee's Ability. In: Boston C, Rudner L, editors. The Basics of Item Response Theory: ERIC Clearinghouse on Assessment and Evaluation. pp. 85-102.  
 Baker FB, Kim S-H (2004a). Bayesian Parameter Estimation Procedures. In: Baker FB, Kim S-H, editors. Item Response Theory Parameter Estimation Techniques, Second Edition. Monticello, New York: Marcel Dekker. pp. 171-201.



- Baker FB, Kim SH (2004b). Parameter Estimation via MMLE and an EM Algorithm. In: Baker FB, Kim S-H, editors. *Item Response Theory Parameter Estimation Techniques*, Second Edition. Monticello, New York: Marcel Dekker. pp. 157-75.
- Bejar I (1983). Introduction to Item Response Models and Their Assumptions. In: Hambleton RK, editor. *Applications of Item Response Theory*. Vancouver, Canada: Educational Research Institute of British Columbia. pp. 1-23.
- Birnbaum A (1968). Some Latent Trait Models. In: Lord F, Novick M, Birnbaum A, editors. *Statistical theories of mental test scores*. Canada: Addison-Wesley Publishing. pp. 397-423.
- Bock RD (2003). Estimation in BILOG-MG. In: Du Toit M, editor. *IRT from SSI: Bilog-MG, multi-log, parscale, testfact*. Lincolnwood, IL: Scientific Software International. pp. 599-610.
- Brown JD (2014). Maximum-Likelihood Estimation. *Linear Models in Matrix Form: A Hands-On Approach for the Behavioral Sciences*. Cham: Springer International Publishing. pp. 69-104.
- Cai L, Choi K, Hansen M, Harrell L (2016). Item Response Theory. *Annu. Rev. Stat. Appl.* 3(1):297-321. doi: 10.1146/annurev-statistics-041715-033702.
- Chen W-H, Thissen D (1999). Estimation of item parameters for the three-parameter logistic model using the marginal likelihood of summed scores. *Br. J. Math. Stat. Psychol.* 52(1):19-37. doi: 10.1348/000711099158946.
- Crocker L, Algina J (1986). Introduction to Item Response Theory. In: Crocker L, Algina J, editors. *Introduction to classical and modern test theory*. Orlando, Florida: ERIC. pp. 339-75.
- Embretson SE, Reise S (2000a). Calibrating Items: Estimation. In: Embretson SE, Reise S, editors. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates. pp. 187-225.
- Embretson SE, Reise S (2000b). Item Response Theory as Model-Based Measurement. In: Embretson SE, Reise S, editors. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates. pp. 40-61.
- Embretson SE, Reise S (2000c). Measuring Persons: Scoring Examinees with IRT Models. In: Embretson SE, Reise S, editors. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates. pp. 158-86.
- Embretson SE, Reise S (2000d). The New Rules of Measurement. In: Embretson SE, Reise S, editors. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates. pp. 13-39.
- Hambleton R, Swaminathan H, Rogers J (1991a). Ability and Item Parameter Estimation. In: Foster D, editor. *Fundamentals of Item Response Theory*. California: Sage Publications. pp. 32-52.
- Hambleton R, Swaminathan H, Rogers J (1991b). Background. In: Foster D, editor. *Fundamentals of Item Response Theory*. California: Sage Publications. pp. 1-6.
- Hambleton R, Swaminathan H, Rogers J (1991c). Concepts, Models and Features. In: Foster D, editor. *Fundamentals of Item Response Theory*. California: Sage Publications. pp. 7-31.
- Hambleton RK (1989). Principles and selected applications of item response theory. In: Linn RL, editor. *Educational measurement*, 3rd ed. New York, NY: Macmillan Publishing. pp. 147-200.
- Hambleton RK, Swaminathan H (1985). Ability Scales. In: Madaus G, Stufflebeam D, editors. *Item Response Theory, Principles and Application*. Norwell, MA: Kluwer Nijhoff Publishing. pp. 53-74.
- Longford N (2008). Maximum Likelihood Estimation. In: Casella G, Fienberg S, Olkin I, editors. *Studying Human Populations: An Advanced Course in Statistics*. New York, NY: Springer New York. pp. 37-66.
- Lord FM (1980a). Estimating ability and item parameters. In: Lord FM, editor. *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers. pp. 179-91.
- Lord FM (1980b). Item Response Theory - Introduction and Preview. In: Lord FM, editor. *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers. pp. 11-26.
- Mahmud J, Sutikno M, Naga DS (2016). Variance difference between maximum likelihood estimation method and expected A posteriori estimation method viewed from number of test items. *Educ. Res. Rev.* 11(16):1579-89. doi: 10.5897/ERR2016.2807.
- Makridakis S, Wheelwright SC, McGee VE (1999). Pengantar. In: Suminto H, editor. *Metode dan aplikasi peramalan*. Jakarta: Binarupa Aksara. pp. 13-25.
- Naga DS (1992a). Estimasi Parameter Pada Model Logistik. In: Naga DS, editor. *Pengantar teori sekor pada pengukuran pendidikan*. Jakarta: Gunadarma. pp. 246-85.
- Naga DS (1992b). Pengertian dan Model Teori Responsi Butir. In: Naga DS, editor. *Pengantar teori sekor pada pengukuran pendidikan*. Jakarta: Gunadarma. pp. 156-92.
- Steinberg L, Thissen D (2013). Item response theory. *The Oxford handbook of research strategies for clinical psychology*. pp. 336-73.
- Swaminathan H (1983). Parameter Estimation in Item Response Models. In: Hambleton RK, editor. *Applications of Item Response Theory*. Vancouver, Canada: Educational Research Institute of British Columbia. pp. 24-44.
- Zanon C, Hutz CS, Yoo H, Hambleton RK (2016). An application of item response theory to psychological test development. *Psicol-Reflex Crit.* 29(1):18. doi: 10.1186/s41155-016-0040-x.