

# The Estimation Stability Comparison of Participants' Abilities on Scientific Literacy Test Using Rasch and One- Parameter Logistic Model

*by* Susongko Purwo 21

---

**Submission date:** 11-Apr-2022 01:20PM (UTC+0700)

**Submission ID:** 1807561336

**File name:** The Estimation Stability Comparison of Participants' Abilities on Scientific Literacy Test Using Rasch and One-Parameter Logistic Model.pdf (548.96K)

**Word count:** 3778

**Character count:** 19968

PAPER • OPEN ACCESS

3  
The Estimation Stability Comparison of  
Participants' Abilities on Scientific Literacy Test  
Using Rasch and One-Parameter Logistic Model

To cite this article: P Susongko 2021 *J. Phys.: Conf. Ser.* **1842** 012037

10  
View the [article online](#) for updates and enhancements.

You may also like

- 4
- [The profile of high school students' scientific literacy on fluid dynamics](#)  
Parno, L Yuliati and N Munfaridah
  - [Scientific and quantitative literacy: a comparative study between STEM and non-STEM undergraduates taking physics](#)  
B Uzpen, A K Houseal, T F Slater et al.
  - [Profile of students' scientific literacy on particle dynamics](#)  
S Mahtari, M Wati, S Rizky et al.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

**Extended abstract submission deadline: April 22, 2022**

Connect. Engage. Champion. Empower. Accelerate.

**MOVE SCIENCE FORWARD**



Submit your abstract



## The Estimation Stability Comparison of Participants' Abilities on Scientific Literacy Test Using Rasch and One-Parameter Logistic Model

P Susongko \*

Science Education Department, Universitas Pancasakti Tegal

\*purwosusongko@upstegal.ac.id

**Abstract.** This research aims at comparing the estimation stability of participants' abilities on scientific literacy test based on the integrated Sciences using Rasch Dichotomous Model and One-parameter Logistic Model (1-PL IRT). The research data was the responses of scientific literacy test based on the integrated Sciences involving 310 students joining the Mathematics and Natural Sciences Programs at SMA N 2 and SMA N 3 Tegal City, Indonesia. The estimation of test participants' abilities with those two models was assisted with the eRm package and ltm package contained in the R program version 4.0.2. The next stage included analyzing mean, median, and variance of the estimation scores resulted from both models. Bootstrapping was then conducted to the parameter estimation using the new samples which have the same size with the original samples starting from 500, 800, 1000, 1500, 2000, 2500, 3000 to 4000. The research result showed that the estimation of participants' abilities in scientific literacy test based on the integrated Sciences using the Rasch Dichotomous Model was more stable than that using the 1-PL IRT seen from both bias and error standard. This means that Rasch modeling on high-stakes test assessments is more precise than using the One-Parameter Logistic Model.

### 1. Introduction

Modern psychometric theories used in the recent education and social science assessments are Rasch model and Item response Theory [1]. The use of Rasch model and Item Response theory in high-stakes situation tests has given adequately effective results from the psychometric aspects with some paradigm differences [2]. Item response theory (IRT) generally refers to three probabilistic measurement models consisting of One-parameter Logistic Model (identical with the dichotomic Rasch model), 2- Logistic Model parameter, and 3- Logistic Model parameter named after the number of parameter items estimated in each model. Those three models can be determined from one probabilistic function for the presence of the correct answers by someone to one item. The probability function of someone answering correctly in the 1-PL IRT model is formulated as follows:

$$P(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}} ; i=1,2,3 \dots n \quad (1)$$

Where:

$P_i(\theta)$  = Probability randomly chosen by the examinee with the ability of  $\theta$  answers item  $i$  correctly

$b_i$  = Item  $i$  difficulty parameter

$n$  = number of item in the test



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

D= a scaling factor introduced to make the logistic function as close as possible to the normal o give function, D=1.7 [3, 4, 5]

The simplest Rasch Model is the dichotomous scoring. The dichotomous test items are test items in which the answers consisting of two categories, such as true-false, correct-incorrect, good-bad, and others. The multiple choice test form generally uses the dichotomous scoring. The dichotomous Rasch Model is explained in the following equation 2.

$$P(\theta) = \frac{e^{\theta-b_i}}{1+e^{\theta-b_i}} ; i=1,2,3 \dots n \quad (2) \quad [6,7,8]$$

Rasch Model is basically the 1-PL IRT model with D=1. It shows that the use of Rasch Model does not require the normality assumption either for the participants' abilities or the test items' difficulty levels.

The use of Rasch model according to some experts is considered more objective than that of One-parameter Logistic Model in analyzing the dichotomous test responses. This is caused by some elements including: (1) Rasch modeling does not require the normality assumptions either to distribute the difficulty levels or to distribute the test participants' abilities, (2) the sample minimum size to estimate the parameter in the Rasch modeling is much smaller than the sample size used in One-parameter Logistic Model [9]. The use of One-parameter Logistic Model is basically looking for the most appropriate model with the existing data, so that if the data do not fit with the model, the other model with more parameters will be selected. Meanwhile, in the use of Rasch model, model is the operationalization of the basic measurement requirements. If the data do not fit with the model, the unfitness reason will be investigated as anomaly [10]. In other words, the implementation of One-parameter Logistic Model or item response model is generally descriptive since aiming at adjusting the model with the data. Conversely, Rasch model is prescriptive since emphasizing on the data adjustment to the model [11].

According to Mok and Wright, Rasch model is also considered as the most satisfying model since at least meeting the requirements as an ideal model by fulfilling five criteria: (1) resulting in a linear size with the same interval, (2) precise estimation process, (3) identifying the inappropriate (*misfits*) or extraordinary (*outliers*) items, (4) having the ability to cope with the missing data, (5) resulting in an independent measurement from the parameter under study [12]. In addition, Rasch model is considered having the ability to convert the learning achievement result scores or the other psychological attribute measurements from the ordinal scale to interval scale [13,14,15].

Rasch model as a measurement model in which during this time is considered as the most objective one, especially when compared with the use of 1-PL IRT model or the other IRT models, is necessary to be proven its effectiveness, especially in the parameter estimation precision either for the item parameter or test participant parameter. The parameter estimation precision covers both accuracy and precision aspects. Accuracy shows the close relationship between the estimation result and the real scores, while precision shows how close the score differences when the estimation repetition is performed. One of precision measurements in the test participant parameter estimation is the score stability parameter when the estimation is repeatedly performed with the same samples. This research aims at revealing how far the test participant parameter estimation stability in the use of Rasch model when compared with that of 1-PL IRT model.

## 2. Method

To achieve the research objectives, some stages are performed: (1) determining the empirical data used, (2) estimating the test participants' abilities in the empirical data by using the Rasch model and 1-PL IRT model (3) Determining a method to analyze the test participant parameter estimation stability, and (4) determining the test participant parameter estimation stability indicators.

The empirical data used are the responses of XII Grade Mathematics and Natural Sciences (MIPA) Program students of Senior High School during the science literacy test based on the integrated Sciences held at SMAN 2 and SMAN 3 Tegal City. The test was held on 11 February 2020 and 21 February

2020. This science literacy test was designed as the minimum competency test to the science literacy mastery with the standard of PISA 2015 and presented by analyzing 14 themes of the integrated mathematics and natural sciences given. Each theme consisted of three items that simultaneously the test contained 42 items. Before the test was used, it had been previously tested to 112 test participants. By using the Messick validity and the implementation of Rasch Model, the instruments were considered valid [16]. The test results were used as one requirement for the Senior High School students to graduate from the Mathematics and Natural Sciences Programs of SMAN 2 and SMAN 3 Tegal city. The test objective and design showed that the test is classified into high stakes test. The high stakes test was used to fulfill the assumptions to the utilization of Rasch Model and 1-PL IRT model. 310 students participated in the test and fulfilled the minimum adequacy in the parameter estimation when using the Rasch Model and 1-PL IRT model [17,18,19].

The test participant ability estimation with the Rasch model and 1-PL IRT model each used the eRm and ltm packages available in the R program version 4.0.2. The next stage was determining the descriptive statistic parameters, such as mean, median and variance from both ability score distribution types. To figure out the parameter estimation stability, the resampling-based method known as Bootstrapp method was used. Bootstrapp method is basically a nonparametric technique to simulate a parameter distribution [20,21]. The use of this Bootstrap method can reduce the sampling error impact of error carryover at the score interval estimation [22,23]. Furthermore, the use of Bootstrapp method was due to the parameter distribution assumption differences of both test participant response models in which the test participant parameter in Rasch model did not have a certain distribution, while the 1-PL IRT model required the normal or parametric distribution requirement. The Bootstrapp method resulted in the statistical inferences in the form of error and bias estimation standards, trust interval, and hypothesis test without assumption, such as the normal or same variance distribution requirement. Thus, the bootstrap method can be more accurate than the classical conclusion based on the normal or t distribution [24].

The next stage was performing the bootstrapping to the mean, median, and variance estimation from both score distributions using new samples possessing the size same with the original samples. The total samples used were from 500, 800, 1000, 1500, 2000, 2500, 3000 up to 4000. The estimation stability used the resulted bias and error standard measurement. The bootstrapping analysis was conducted and supported with the SPSS program version 21. The bias showed the mean statistical differences between all *bootstrap* samples and the original samples. The error standard showed the error standard from all mean scores of *bootstrap* samples. Smaller error standard and bias scores in both score distributions (Rasch model and 1-PL IRT Model) showed higher stability.

### 3. Result and Discussion

The data analysis was started by estimating the descriptive statistical parameter of the test participants' abilities using the Rasch model and 1-PL IRT Model. Furthermore, the test participant ability distributions were then estimated based on mean, median, standard deviation, and variance as presented in Table 1.

**Table 1.** Descriptive Statistical Parameters of the participants' abilities on science literacy test based on the integrated mathematics and natural sciences using Rasch Model and 1-PL IRT Model

	Rasch Model	1-PL IRT
N		
Valid	310	310
Missing	0	0
Mean	.1422	-.0216
Median	.1400	.0370
Std. Deviation	.57462	.79363
Variance	.330	.630

Table 1 shows that the mean and median scores estimated using the Rasch Model gave bigger scores than those estimated using the 1-PL IRT Model. Meanwhile, the distribution measurements were



standard deviation and variance estimated using the Rasch Model which gave smaller scores when compared to those estimated using the 1-PL IRT Model. It means that the estimation using the Rasch model gave the deviation scores smaller than the ability mean scores estimated using the 1-PL IRT Model.

To figure out the parameter estimation stability, the resampling was performed that a total of 500, 800, 1000, 1500, 2000, 2500, 3000 up to 4000 samples were obtained with the same size as many as 310 participants and each mean, median and variance parameter was obtained through the bootstrapping analysis. The bias and error standard from the estimation of each sample number can be seen in Table 2 and Table 3.

**Table 2.** The descriptive parameter estimation bias of ability scores obtained through Resampling using the Rasch Model and 1-PL IRT Model

parameter	Model	500	800	1000	1500	2000	2500	3000	4000
Mean	Rasch	.0001	.0008	.0003	-.0014	.0000	.0000	-.0008	.0000
	1-PL	.0024	-.0002	.0010	.0006	-.0011	.0004	-.0005	.0003
Median	Rasch	-.0007	-.0008	-.0018	-.0018	-.0020	-.0010	-.00145	-.00178
	1-PL	-.0597	-.0670	-.0592	-.0585	-.0628	-.0617	-.00181	-.00175
Variance	Rasch	-.003	-.001	.000	-.001	-.002	-.001	-.001	-.001
	1-PL	-.001	-.002	-.004	-.004	-.004	-.001	-.002	-.002

Table 2 shows that all sample number variations indicated that the mean, median, and variance estimation bias of test participants' abilities in the use of Rasch Model were smaller than those in the use of 1-PL IRT Model. Moreover, in the sample number of 2000, 2500, and 4000, for the mean parameter estimation had the bias reaching 0.0000 or no bias. Meanwhile, for the variance parameter estimation, in the sample number of 1000 had the bias of 0.000 or no bias. The mean parameter estimation has smaller bias than the median and variance in all sample numbers used. The analysis results showed that the test participants' abilities estimated using the Rasch Model were more stable than using the 1-PL IRT Model from the bias aspect.

**Table 3.** Error Standard of Descriptive Parameter Estimation on the ability Scores Using Resampling in the use of Rasch Model and 1-PL IRT Model

parameter	Model	500	800	1000	1500	2000	2500	3000	4000
Mean	Rasch	.0351	.0331	.0337	.0352	.0337	.0335	.0336	.0337
	1-PL	.0478	.0442	.0460	.0449	.0463	.0460	.0449	.0468
Median	Rasch	.0279	.0265	.0286	.0278	.0281	.0268	.0273	.0307
	1-PL	.0843	.0819	.0808	.0817	.0826	.0833	.0817	.0825
Variance	Rasch	.029	.028	.028	.029	.028	.029	.029	.029
	1-PL	.047	.048	.048	.047	.048	.049	.048	.048

Table 3 shows that all sample number variants indicated that error standard as well as mean, median and variance estimation from the test participant's abilities in the use of Rasch Model were smaller than those in the use of 1-PL IRT Model. The analysis results showed that the test participant ability estimation using the Rasch Model was more stable than that using the 1-PL IRT Model from the error standard aspect.

The research result showed that the test participant parameter estimation using the Rasch Model was more stable than that using the 1-PL IRT Model or it could be said that the use of Rasch Model has higher precision in estimating the test participant parameter from the precision aspect. The research

results also supported the excellence of Rasch model in which during this time is considered the most objective one. However, the excellence of a model to the responses of test participants was not only viewed from the objectivity aspect but also seen from the ability aspect in scoring [25]. The weakness of scoring using the Rasch model has not revealed the test participant response patterns. The 2-PL IRT and 3-PL IRT model can do more comprehensive scoring based on the test participant response patterns [17]. Although the use of 1-PL IRT Model is similar with Rasch Model, that is, unable to reveal the test participant response patterns, this research proved that Rasch Model had higher precision level in estimating the test participant ability parameter when compared to that using the 1-PL IRT Model.

The previous studies showed that the use of Rasch model and 1-PL IRT Model had paradigmatic differences [2]. Although both Rasch model and 1-PL IRT Model are item responsive models with similarities in their mathematical structural models used, both used a social measurement approach from different start points. The main different is that IRT prioritizes more on data and aims at revealing the most satisfying item responsive model in explaining the data, while Rasch model prioritizes items intended to find items best matched with the data. The other different principle is that Rasch model describes the interaction between variables connected with a relationship pattern, while IRT model explains the related probability distribution with the item responses in the psychometric test as the latent variable function [4]. Those differences should be well understood that the researchers can select the best approach to use in accordance with the research objectives.

The benefits of both models (Rasch model and 1-PL IRT Model) are in making the estimation which can be truly free from samples since both only use one parameter that both can meet the requirements as an objective measurement and do the conversion from the ordinal score to the interval score [28,29]. The weakness of both models when compared to the other IRT models is related to their ability in estimating the test participants' ability which is still based on the total score instead of the response pattern [30, 31].

Not many previous studies compared the use of Rasch Model with 1-PL IRT Model. The researchers more frequently compared the use of classical test theories with the use of Rasch model or IRT Model. Various studies showed that the use of Rasch model and IRT Model (1PL, 2PL,3PL) in the educational and psychological measurements provide more precise results when compared to the use of classical test theories [32,33,34]

#### 4. Conclusion

The ability of participants joining the scientific literacy test based on the integrated mathematics and natural sciences estimated using the Rasch Dichotomous Model is more stable when compared to that using the One-parameter Logistic Model (1-PL IRT) seen from both bias and error standard. Thus, it can be concluded that Rasch Model has higher precision level in the test participant ability parameter estimation when compared to that using the 1-PL IRT Model. This means that Rasch modeling on high-stakes test assessments is more precise than using the One-Parameter Logistic Model.

#### References

- [1] Kean J, Brodke D S, Biber J, & Gross P 2018 Brain impairment: a multidisciplinary journal of the Australian Society for the Study of Brain Impairment **19** 1 91.
- [2] Petrillo J, Cano SJ, McLeod LD, & Coon C D 2015 Value in Health **18** 1 25-34.
- [3] Van der Linden W J 2018 Handbook of item response theory three volume set CRC Press.
- [4] Marsman M, Borsboom D, Kruis J, Epskamp S., van Bork R, Waldorp L J & Maris G 2018 Multivariate behavioral research **53** 1 15-35
- [5] Scott T F & Schumayer D 2015 Physical Review Special Topics-Physics Education Research **11** 2 020134.
- [6] Bond TG & Fox CM 2015 Applying the Rasch model: fundamental measurement in the human sciences (Vol. 1) New York: Routledge
- [7] Goh H E, Marais I & Ireland MJ 2017 Assessment **24** 3 387-398.
- [8] Sousa L B, Prieto G, Vilar M, Firmino H, & Simões MR 2015 Research on aging **37** 8

- 787-814.
- [9] De Ayala RJ 2018 Item Response Theory and Rasch Modeling. The Reviewer's Guide to Quantitative Methods in the Social Sciences, 145-163.
  - [10] Andrich D, & Marais I 2019 Comparisons and Contrasts Between Item Response Theory and Rasch Measurement Theory. In A Course in Rasch Measurement Theory (pp. 221-229). Springer, Singapore.
  - [11] Chong Ho Yu 2017 A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling retrieved from Website: <https://creative-wisdom.com/computer/sas/IRT.pdf>
  - [12] Sumintono, B. 2018 Rasch Model Measurements as Tools in Assesment for Learning. In 1st International Conference on Education Innovation (ICEI 2017). Atlantis Press.
  - [13] Zhao Y, Huen J M, & Chan YW 2017 Research in Higher Education **58** 6 605-616.
  - [14] Medvedev O N, Siegert R J, Feng X J, Billington DR, Jang JY, & Krägeloh CU 2016 Mindfulness **7** 2 384-395
  - [15] Cupani M, Morán V E, Ghío FB, Azpilicueta A E, & Garrido SJ 2020 Journal of Child and Family Studies **29** 9 2472-2486.
  - [16] Ravand H, & Firoozi T 2016 International Journal of Language Testing **6** 1 1-18
  - [17] Custer M 2015 Sample Size and Item Parameter Estimation Precision When Utilizing the One-Parameter" Rasch" Model. Online Submission
  - [18] Tay L, Meade AW, & Cao M 2015 Organizational Research Methods **18** 1 3-46.
  - [19] Uyigüe AV & Orheruata MU 2019 Test Length and Sample Size for Item-Difficulty Parameter Estimation in Item Response Theory.
  - [20] Saha M, Dey S, & Maiti SS 2018 Journal of Industrial and Production Engineering **35** 3 160-169.
  - [21] Dosne AG, Bergstrand M, Harling K, & Karlsson M O 2016 Journal of pharmacokinetics and pharmacodynamics **43** 6 583-596.
  - [22] Liu Y, & Yang J S 2018 psychometrika, **83** 2 333-354.
  - [23] Roodman D, Nielsen MØ, MacKinnon JG & Webb MD 2019 The Stata Journal **19** 1 4-60.
  - [24] Hesterberg T C 2015 The American Statistician **69** 4 371-386.
  - [25] Lemoine F, Entfellner JBD, Wilkinson E, Correia D, Felipe MD, De Oliveira, T & Gascuel O 2018 Nature **556** 7702 452-456.
  - [26] Preti A., Vellante M, & Petretto DR 2017 Cognitive Neuropsychiatry **22** 3 233-253.
  - [27] OLMUŞ H., Nazman E., & Erbaş S 2017 Gazi University Journal of Science **30** 1
  - [28] Medvedev O N, Krägeloh CU, Hill E M., Billington R, Siegert R J, Webster CS & Henning MA 2019 Journal of Health Psychology **24** 8 1070-1081.
  - [29] Planinic M., Boone W J, Susac A & Ivanjek L 2019 Physical Review Physics Education Research, **15** 2 020111.
  - [30] Lacourly N, Martin J S., Silva M, & Uribe P 2018 IRT scoring and the principle of consistent order. arXiv preprint arXiv:1805.00874.
  - [31] Żółtak T & Golonka G 2015 Edukacja, **3**, 63-76.
  - [32] Himelfarb I 2019 Journal of Chiropractic Education **33** 2 151-163.
  - [33] Mallinckrodt B, Miles JR. & Recabarren DA 2016 The Counseling Psychologist **44** 2 146-194.
  - [34] Bourion-Bédès S, Schwan R., Epstein J., Laprevote V, Bédès A., Bonnet J L, & Baumann, C 2015 Quality of Life Research **24** 2 287-293.
  - [35] Kohli N, Koran J, & Henn L 2015 Educational and Psychological Measurement **75** 3 389-405.
  - [36] Bergan JR 2013 Rasch versus Birnbaum: New arguments in an old debate. Tucson, AZ: Assessment Technology.
  - [37] Gordon R A 2015 Journal of Marriage and Family **77** 1 147-176.
  - [38] von Davier M & von Davier AA 2013 Journal of Educational Measurement **50** 3 295-303.



# The Estimation Stability Comparison of Participants' Abilities on Scientific Literacy Test Using Rasch and One-Parameter Logistic Model

## ORIGINALITY REPORT

14%

SIMILARITY INDEX

13%

INTERNET SOURCES

10%

PUBLICATIONS

%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://www.researchgate.net">www.researchgate.net</a> Internet Source	4%
2	<a href="http://repository.ung.ac.id">repository.ung.ac.id</a> Internet Source	2%
3	<a href="http://www.sciencegate.app">www.sciencegate.app</a> Internet Source	1%
4	R O Diyanti, S Mulyani, Sarwanto. "Identify students' scientific literacy through tests on kinetics of chemistry concept", Journal of Physics: Conference Series, 2020 Publication	1%
5	<a href="http://eudl.eu">eudl.eu</a> Internet Source	1%
6	<a href="http://pdfs.semanticscholar.org">pdfs.semanticscholar.org</a> Internet Source	1%
7	<a href="http://journals.sagepub.com">journals.sagepub.com</a> Internet Source	1%

8	"Rasch Measurement", Springer Science and Business Media LLC, 2020 Publication	1 %
9	J Hobart, S Cano. "Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods", Health Technology Assessment, 2009 Publication	1 %
10	<a href="http://www.coursehero.com">www.coursehero.com</a> Internet Source	1 %
11	Siswiyanti, Rusnoto, Saufik Luthfianto, Nurjanah, Arif Zainudin. "The Work Posture Assessment Towards Musculoskeletal Disorders in Coloring Activities of Indonesian Hand Drawn Batik", Journal of Physics: Conference Series, 2019 Publication	<1 %
12	Vahid Aryadoust, Li Ying Ng, Hiroki Sayama. "A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research", Language Testing, 2020 Publication	<1 %
13	<a href="http://etheses.dur.ac.uk">etheses.dur.ac.uk</a> Internet Source	<1 %
14	<a href="http://hdl.handle.net">hdl.handle.net</a> Internet Source	<1 %

15

Black, R. A., and S. F. Butler. "Using the GLIMMIX Procedure in SAS 9.3 to Fit a Standard Dichotomous Rasch and Hierarchical 1-PL IRT Model", Applied Psychological Measurement, 2012.

Publication

<1 %

---

Exclude quotes Off

Exclude matches Off

Exclude bibliography On