International Journal of Education and Humanities (IJEH), 1(4) 2021:171-180
http://i-jeh.com/index.php/ijeh/index
e-ISSN :2798-5768

Comparative Analysis of Question Item Parameters and Students' Ability between Dichotomy and Polytomic Score Versions; Research on Mathematics National Exam Test Participants

Reza Oktiana Akbar [1]

Abstract

The purpose of this study was to determine the differences in the estimation of item parameters and students' abilities between the dichotomous scoring version and the polytomy scoring version. The data used in this study is the response of participants in a particular year national mathematics exam test which is presented in two versions of the scoring form, namely dichotomy and polytomy. In the dichotomous scoring version, each item is scored individually. Alternatively, each item that is scored separately on the dichotomous version, is aggregated into the same indicator, and then summed to obtain the polytomy version. Estimation of item parameters and students' abilities in the dichotomous scoring version was carried out using the 3 PL (Logistics Parameter) model using MLE estimation and the polytomy scoring version using the GPCM (Generalized Partial Credit Model) using EAP. Both were analyzed using PARSCALE software. Comparative analysis of the two scoring versions by looking at the average results of the estimated difficulty level, graph analysis, calculating correlations, and the results of the value of the information function. The results of the analysis show that the average difficulty level of the dichotomous version is 0.166, with a standard deviation of 1.137, while the average difficulty level of the polytomy version is 0.033, with a standard deviation of 0.940. The value of the dichotomous scoring version of the information function is higher than the polytomy scoring version. These results indicate that the math exam test with the dichotomous scoring version is better than the polytomy scoring version. graph analysis, calculate correlations, and result in value information functions. The results of the analysis show that the average difficulty level of the dichotomous version is 0.166, with a standard deviation of 1.137, while the average difficulty level of the polytomy version is 0.033, with a standard deviation of 0.940. The value of the dichotomous scoring version of the information function is higher than the polytomy scoring version.

Keywords: Student's Ability, Dichotomy Scoring, Politomy Scoring, MLE, EAP.

A. Introduction

Often national exam questions, especially mathematics subjects, usually use a scoring technique in the form of a dichotomy, but other scoring alternatives can also be used. The scoring alternative can be made into a polytomy form by collecting each question in the

[1] IAIN Syekh Nurjati Cirebon, Indonesia, West Java, Indonesia, rezaoktiana@syekhnurjati.ac.id

dichotomous scoring version into the same question indicator. The exact estimation of ability given in the two versions of the scoring form, both dichotomy and polytomy may also be different, but this still needs to be studied further. This study only focuses on the comparison of students' ability estimates using a dichotomous score and students' ability estimates using a polytomy score.

An item analysis to determine item characteristics and test-taking student ability estimation can be done using 2 approaches, namely the item response theory of the test using the 3-parameter logistic model (3PL) for the dichotomous score and the Graded Partial Credit Model for the polytomy score. According to Keeves and Alagumalai(in Retnawati, 2015), item response theory approach is an alternative approach that can be used in analyzing a test. There are two principles used in this approach, namely the principle of relativity and the principle of probability.

Based on this principle, a logistic model can be drawn up by connecting a person's probability of answering correctly with a scale of ability (θ), difficulty level (b), item discriminating power (a), and pseudo guessing (c). If the probability is expressed by P, the ability with and the level of difficulty with b, then the relationship between the four quantities is expressed by the equation(Hambleton, Swaminathan, & Rogers, 1991). The equation, which is a model containing 3 item parameters, which is then called the 3PL model mathematically can be stated as follows:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Information:

θ : test taker's ability level
$P_i(\theta)$: probability of the test taker having the ability θ can answer item i correctly
a_i : discriminatory index
b_i : item difficulty index i
c_i : i-item pseudo-guess index
e : natural number whose value is close to 2.718
n : the number of items in the test
D : a scaling factor that costs 1.7

Equation model (1), which states the relationship between the probability of occurrence of a phenomenon and the ability that is most often used in psychometry, health, and education. There are two ability estimation methods that can be used, namely the maximum likelihood method and the Bayes method (in Retnawati, 2015, p.146). Related to this, in this study the estimation of dichotomous scoring uses the Maximum Likelihood Estimation (MLE) method.

According to Retnawati (2015, p.3), the Maximum Likelihood method of the participants' ability scores is estimated by maximizing the function

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

With a function that matches item j.

Next, the implicit likelihood . equation is solved

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

Bayes estimation is the mean of the posterior distribution of, after being given the response pattern of the participants on the x_i test results, according to can be approximated accurately by the equation:

$$\bar{\theta}_i = \frac{\sum_{k=1}^{q} x_k P(x_i|x_k) A(x_k)}{\sum_{k=1}^{q} P(x_i|x_k) A(x_k)}$$

The function of the response pattern x_i is often called the estimator of Expected a Posterior (EAP). The ability estimate given by EAP includes all types of response patterns and also has a smaller mean error than other approaches, including the Maximum Likelihood approach. Although the sample mean of the EAP estimate does not provide a biased value for the population mean, generally the sample standard deviation is smaller than the population standard deviation. In practice this is not very influential because the standard deviation of the sample will approach the population if the score is standardized.

At the beginning of the development of the polytomy item response theory, the better known model was an extension of the Rasch model called the Partial Credit Model (PCM). PCM is a polytomy scoring model which is an extension of the Rasch model on dichotomy data. The assumption in PCM is that each item has the same different power. PCM has similarities with the Graded Response Model (GRM) in that the items are scored in a tiered category, but the difficulty index in each step does not need to be ordered, one step can be more difficult than the next step.(Retnawati, 2018).

The general form of PCM according to Muraki & Bock (1997:16) in (Retnawati, 2018) as follows:

$$P_{jk}(\theta) = \frac{exp \sum_{v=0}^{k} (\theta - b_{jv})}{\sum_{h=0}^{m} exp \sum_{v=0}^{k} (\theta - b_{jv})}, k = 0,1,2, \dots, m$$

with:

$P_{jk}(\theta)$:The probability of participants having the ability to get a category k score on item j,
θ : participant's ability
m+1 : number of item categories j,
b_{jv} : index of difficulty category k item j

$$\sum_{h=0}^{k}\left(\theta - b_{jh}\right) \equiv 0 \; dan \; \sum_{h=0}^{k}(\theta - b_{jh}) \equiv \sum_{h=1}^{k}(\theta - b_{jh})$$

Scorethe category on the PCM indicates the number of steps to correctly complete the item. Higher category scores indicate greater ability than lower category scores. In PCM, if an item has two categories, then equation 2 becomes the equation of the Rasch model, as stated by Hambleton, Swaminathan (1985), and also strengthened by Hambleton, Swaminathan, and Roger (1991). As a result of this, PCM can be applied to the items of polytomies and dichotomies.

A further development of polytomy scoring is the Generalized Partial Credit Model (GPCM). According to Muraki (1997), GPCM is a general form of PCM, which is expressed in mathematical form, which is referred to as the response function of item categories as follows.

$$P_{jh}(\theta) = \frac{exp \sum_{v=0}^{k} Z_{jr}(\theta)}{\sum_{h=0}^{m} exp\left[\sum_{v=0}^{k} Z_{jr}(\theta)\right]}, k = 0,1,2, \dots, m_j$$

and

$Z_{jh}(\theta)=D_{aj}(\theta - b_{jh})=D_{aj}(\theta - b_j + d_h)$, $b_{j0}=0$

with

$P_{jk}(\theta)$: the probability of the participant being able θ get a category k score on item j,
θ : participant ability,
a_j : j item discrepancy index,
b_{jh} : difficulty index category k item j,
b_j : item location difficulty index j (location item parameter)
d_k : category parameter k,
m_{j+1} : number of item categories j, and
D : scale factor (D=1.7)

If $\theta = b_{jk}$, then $P_{jk}(\theta) = P_{jk}\text{-}1(\theta)$

If $\theta > b_{jk}$, then $P_{jk}(\theta) > P_{jk}\text{-}1(\theta)$

If $\theta < b_{jk}$, then $P_{jk}(\theta) < P_{jk}\text{-}1(\theta)$, K=1,2,3,...,$m_j$

GPCM is formulated based on the assumption that every probability of choosing the kth category beyond the (k-1) category is built by the dichotomy model. Pjk is a special probability of choosing the kth category from m_j +1 categories. The relationship between the probability of answering correctly for each ability θ presented in a Categorical Response Function (CRF) graph (du Toit, 2003).

B. Methods

Types of research

This research is an exploratory descriptive study with the aim of comparing the estimation of students' abilities by using dichotomous and polytomy scoring on the UN questions based on the results of the responses of junior high school students in DIY in the 2013/2014 academic year.

Research Subjects and Objects

The research subjects were junior high school students in Yogyakarta who took the Mathematics National Examination for the 2013/2014 academic year. In this study, 3000 test takers took the test results for analysis. The objects selected were multiple-choice objective mathematics questions consisting of 40 items with a dichotomy score and 25 new items with a polytomy score.

Data collection technique

In this study, the technique used in collecting data is documentation technique, by collecting student responses to the mathematics national exam questions in Yogyakarta as many as 46422 students and the data analyzed is as many as 3000 students.

Data analysis technique

Analysis of the mathematical data of the National Examination using two approaches to estimating students' abilities, namely using dichotomous scoring with MLE (Maximum Likelihood) estimation using the PARSCALE program and estimating students' abilities using polytomy scoring using the PARSCALE program. Estimating students' abilities in the second approach uses the Bayesian (EAP) method of estimating students' abilities.

Scoring is dichotomous by giving a score of 1 for correct items and 0 for incorrect items, while for polytomy scoring, the grouping stage is carried out by determining groups of items that have the same/similar question indicators. So that new items are obtained from the grouping of UN items and the results of the grouping carried out are as follows:

No Item	Merging Questions	Score range
1	1	0-1
2	2	0-1
3	3, 4, 5	0-3

4	6	0-1
5	7, 8, 9	0-3
6	10	0-1
7	11, 12	0-2
8	13, 14	0-2
9	15	0-1
10	16, 17, 18	0-3
11	19, 20	0-2
12	21	0-1
13	22	0-1
14	23	0-1
15	24, 25, 26	0-3
16	27	0-1
17	28	0-1
18	29	0-1
19	30, 31	0-2
20	32	0-1
21	33	0-1
22	34, 35	0-2
23	36, 37	0-2
24	38, 39	0-2
25	40	0-1

C. Findings and Discussion

The use of the 3PL model with MLE estimation carried out using Parscale Software estimates 40 items for the national mathematics exam. The estimation results are presented in Table 1 below.

Table1. Parameter 40 Items in Dichotomous Scoring Version

Items	a	b	Items	a	b	Items	a	b	Items	a	b
1	1,295	-0.694	11	1,130	-0.344	21	0.002	0.000	31	1.032	-0.366
2	0.824	-0.412	12	1,330	-0.096	22	0.283	2.033	32	0.689	-1,824
3	1.218	-0.765	13	0.729	1.0692	23	0.735	-0.550	33	1.037	-0.351
4	1,229	-0.344	14	0.950	-0.118	24	0.140	0.742	34	0.774	-0.056
5	0.997	-0.550	15	1,296	-0.329	25	0.553	0.932	35	0.526	1.052
6	1,226	-0.230	16	0.295	0.965	26	0.595	0.111	36	0.755	-0.095
7	1,247	-0.608	17	0.512	0.972	27	1,269	-0.310	37	0.406	0.504
8	1.310	-0.328	18	0.582	0.121	28	0.837	-0.855	38	0.729	-0.454
9	0.621	-0.289	19	0.851	0.068	29	0.672	0.337	39	0.477	-1.046
10	1,153	-0.354	20	1,191	-0.433	30	1.413	-0.399	40	0.754	-0.588

Based on the results in Table 1, it can be seen that almost all items are in the medium category and there is one easy item (number 22). From Table 1, it is also found that there is one very bad question (number 21), because almost all students ignore working on this question (omit) or if it is done carelessly.

The use of the GPCM model with EAP estimation carried out using Parscale Software estimates 25 indicators for the national mathematics exam test questions. The estimation results are presented in Table 2 below.

Table2. Item Indicator Parameters in Polytomy Scoring Version

Items	a	b	Items	a	b	Items	a	b
1	1.022	-0.653	11	0.568	-0.733	21	0.814	-0.195
2	0.755	-0.495	12	1.253	-0.755	22	0.665	0.264
3	0.432	0.470	13	0.763	-0.479	23	0.641	-1,993
4	0.402	0.392	14	1.174	-0.266	24	1.021	-0.391
5	1.217	-0.257	15	1.111	-0.401	25	0.759	-0.606
6	0.801	0.342	16	1,258	-0.368			
7	0.964	-0.239	17	0.043	0.000			
8	1,100	-0.448	18	0.271	2,182			
9	0.626	0.358	19	0.758	0.038			
10	0.550	-0.389	20	1,227	-0.582			

Based on the item parameters, it is possible to draw a plot of the characteristic curve matrix of all items for the dichotomous scoring version. All of these characteristic curves are presented in Figure 1 below.

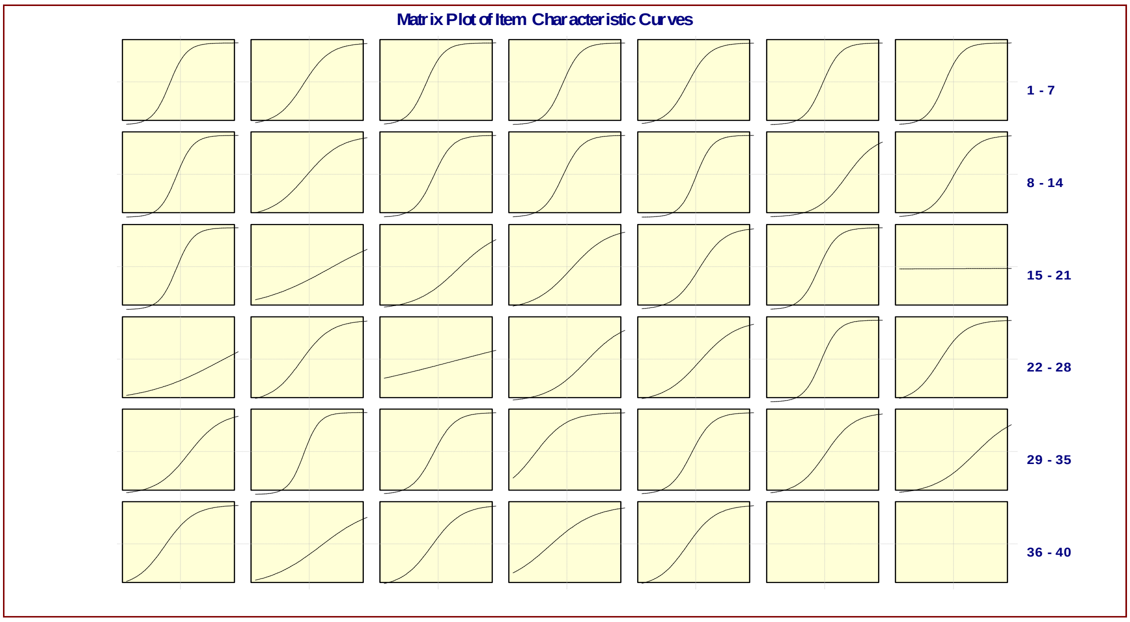

Figure 1. Graph of Category Response Curve (CRC) for National Examination for Dichotomy Data

Based on Figure 1, it is found that almost every item with the same indicator has almost the same pattern tendency. This means that one's ability should be the same when faced with questions with the same indicators, especially if the level of difficulty is the same. For example, the third indicator containing item numbers 3, 4, 5, all three have a level of slope, slope and intersection with almost the same axis. Likewise the 5th, 7th, 8th, 10th, 11th, 19th, 22nd, 23rd and 24th indicators. There is only one indicator in the 15th indicator (consisting of items 24, 25 and 26) which is quite different, namely item 24, the level of slope and slope is significantly different with items 25 and 26.

Based on the item parameters, it is possible to draw a plot of the characteristic curve matrix of all items for the polytomy scoring version. All of these characteristic curves are presented in Figure 2 below.

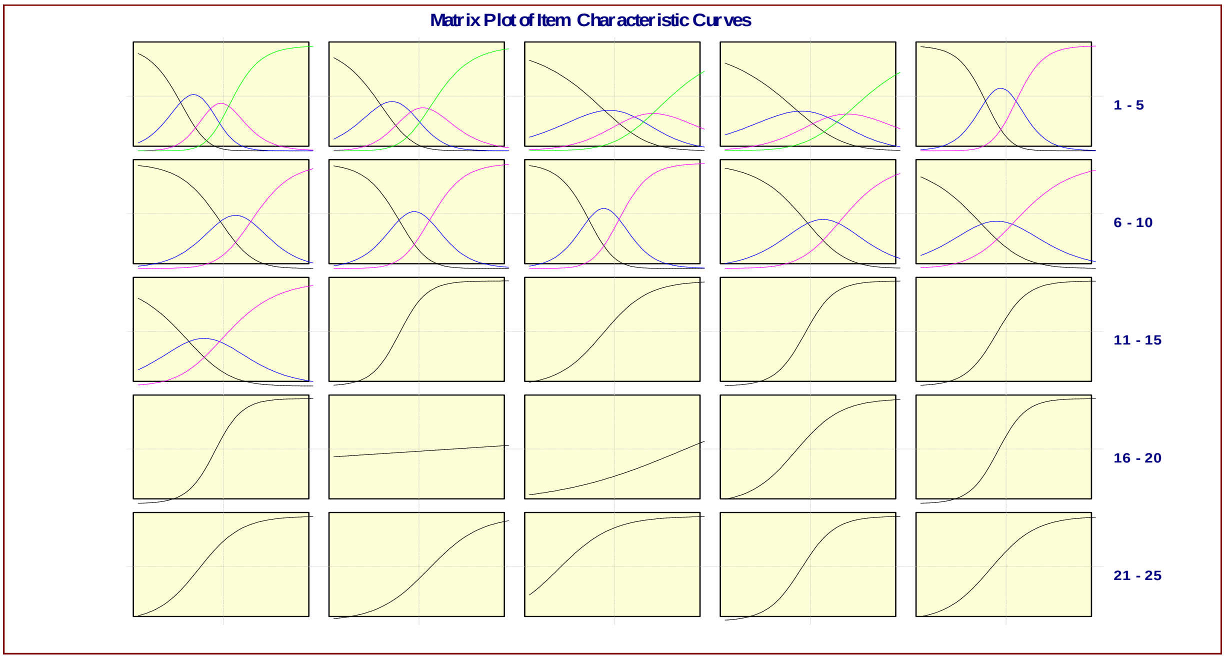

Figure 2. Graph of Category Response Curve (CRC) for National Examination for Politomy Data

Based on Figure 2 and Table 2 it is known that all items are suitable for the Generalized Partial Credit Model (GPCM), but if you pay attention to each category of difficulty index of each item, it can be seen that item number 18 has an easy difficulty index category because it has a value of >2. From Table 2, it is also found that there is one very bad question (number 17), because almost all students ignore working on this question (omit) or if it is done carelessly.

The results of the estimation of students' abilities in the dichotomous and polytomy scoring versions are presented in Table 3 below.

Table 3. Comparison of Mean and Standard Deviation Based on Scoring Version

	Dichotomy	Politomy
Average	0.166	0.033
Stdev	1.137	0.940

Based on the results in Table 4, it is found that the estimation results in the dichotomous scoring version are higher than the polytomy scoring version. Taking the standard deviation into account, the results of the dichotomous version are more variable than the polytomy scoring.

The value of the information function (VIF) can be estimated using the parameters on each indicator item and then the estimation results are concluded. The standard error of measurement can also be estimated using VIF. In the test, the results of VIF and SEM can be presented in Figure 4 (in the dichotomous version) and Figure 5 (in the polytomic version).

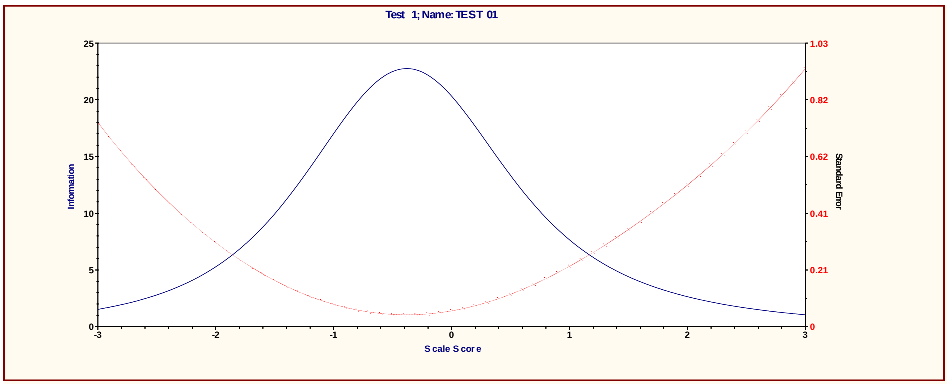

Figure 4. VIF and SEM on Test 1 (Dichotomous Version)

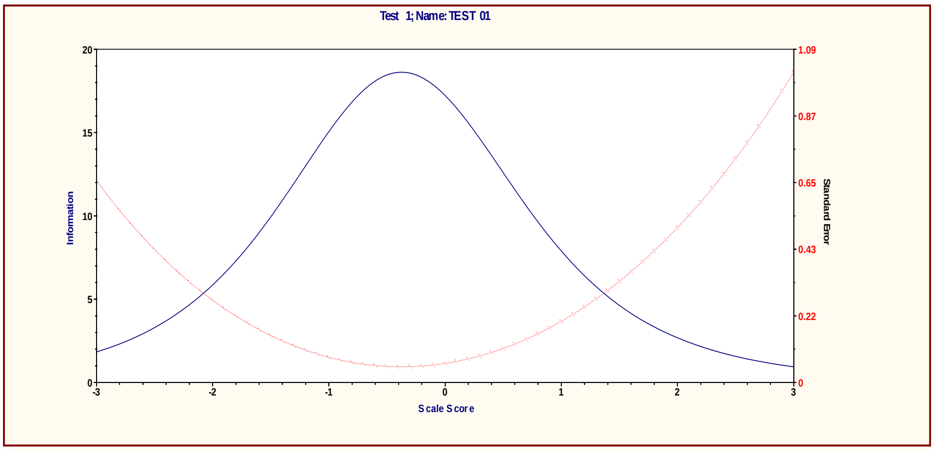

Figure 5. VIF and SEM on Test 1 (Polytomy Version)

In Figure 4, it is shown that the maximum value of the information function is 23 on an ability scale equal to -0.4 . In Figure 5, it is shown that the maximum value of the information function is 19 on an ability scale equal to -0.4. So, this indicates that the value of the information function in the dichotomous scoring version is higher than the value of the information function in the polytomy scoring version.

D. Conclusion

Based on the results of the ability estimation using the dichotomous and polytomy scoring versions, it is concluded that in general the results of the estimation of students' abilities for dichotomous and polytomy scoring do not show significant differences. However, the value of the information function using the dichotomous version with MLE estimation is better than the value of the information function using the polytomy version with GPCM estimation.

References

Du Toit, M. (Ed.). (2003). IRT from SSI: Bilog-MG, multilog, parscale, testfact. Scientific Software International.

Hambleton, R. K., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands.

Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). Fundamentals of Item Response Theory.

Muraki, E. (1997). A generalized partial credit model. In Handbook of modern item response theory (pp. 153-164). Springer, New York, NY.

Retnawati, H. (2015). Perbandingan Estimasi Kemampuan Laten Antara Metode Maksimum Likelihood Dan Metode Bayes. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(2), 145–155. https://doi.org/10.21831/pep.v19i2.5575

Retnawati, H. (2018). Mengestimasi Kemampuan Peserta Tes Uraian Matematika dengan Pendekatan Teori Respons Butir dengan Penskoran Politomi dengan Generalized Partial Credit Model.

Du Toit, M. (Ed.). (2003). IRT from SSI: Bilog-MG, multilog, parscale, testfact. Scientific Software International.

Hambleton, R. K., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands.

Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). Fundamentals of Item Response Theory.

Muraki, E. (1997). A generalized partial credit model. In Handbook of modern item response theory (pp. 153-164). Springer, New York, NY.

Retnawati, H. (2018). Mengestimasi Kemampuan Peserta Tes Uraian Matematika dengan Pendekatan Teori Respons Butir dengan Penskoran Politomi dengan Generalized Partial Credit Model.