Journal Visipena
Volume 16, Number 1, 2025 pp. 16-27
P-ISSN 2086-1397 E-ISSN: 2502-6860
Open Access: https://ejournal.bbg.ac.id/visipena

ACCURACY OF MATHEMATICAL CRITICAL THINKING SKILLS ASSESSMENT WITH A MODERN APPROACH: *GENERALIZED PARTIAL CREDIT MODEL*

Muziyati Hanim[1], Mujahidawati[2], Ilham Falani*[3], Aliffia Teja Prasasty[4], Rina Nurhidayati[5]
[1,2,3]*Master of Mathematics Education, University of Jambi*
[4,5]*Industrial Engineering, Indraprasta PGRI University, Jakarta*

* Corresponding Author: *ilhamfalani@unja.ac.id*

ARTICLEINFO

Article history:
Received : Apr 05, 2025
Revised : May 15, 2025
Accepted : Jun 27, 2025
Available online : Jun 30, 2025

Keywords:
Generalized Partial Credit Model, Mathematical Critical Thinking Skills, Assessment

ABSTRACT

Critical thinking skills mathematical critical thinking skills are essential to face the challenges of the 21st century. These skills in schools is still not optimal because it is not supported by valid and reliable instruments. Instruments that are valid and reliable. This study aims to measure mathematical critical thinking skills through the Generalized Partial Credit Model (GPCM) approach, to increase the accuracy of the assessment. The test instrument used is an indicator-based description test FRISCO (Focus, Reason, Inference, Situation, Clarity, and Overview). The quantitative research method with an experimental design was applied to seventh grade students in four public junior high schools, North Bahar District, Muaro Jambi. North Bahar, Muaro Jambi. Data analysis using GPCM with software PARSCALE 4.1 SOFTWARE. The p value of the fit test of 0.335 (> 0.05) indicates that the question fit the GPCM model. Student ability estimates ranged from -2.23 to 2.35, with the majority in the low ability category. Item parameters showed discrimination values of 0.855-1.534 and difficulty levels of -0.123 to 0.238, indicating good quality and balanced questions. Test information function showed that the instrument effectively measured students' abilities at various levels. These results prove that GPCM is appropriate for developing a valid, valid, and balanced mathematical critical thinking assessment instrument valid, accurate, and informative.

INTRODUCTION

In an era of rapid change and relentless innovation, 21st century skills are key to meeting global challenges and opportunities. These skills include the ability to communicate, collaborate, think critically, creatively, and be technologically and information literate (Atiaturrahmaniah *et al*., 2022; Tohani & Aulia, 2022). In a world that is constantly changing, 21st century skills are the main provision for a person to be able to adapt, solve complex problems, and compete productively in various areas of life (Mardhiyah *et al*., 2021).

One of the most crucial 21st century skills is critical thinking. Critical thinking is the intellectual ability to actively and skilfully evaluate and process information to make decisions or believe in something (Adhikari, 2023; Cahayu *et al*., 2024). Mathematics not only trains logical and numerical abilities, but also cultivates analytical, evaluative, and problem-solving abilities that are all key elements of critical thinking (Mursidah *et al*., 2023).

Mathematical critical thinking allows students to analyze information, evaluate solutions, formulate logical arguments, and identify patterns or assumptions in solving mathematical problems. Students who have these skills will be better prepared to face challenges both in academic contexts and in daily life. However, for the development of these skills to be optimal, proper and accurate assessments are needed (Zalukhu *et al*., 2023).

Assessment is a process used to assess student learning outcomes based on data that has been collected through measurement. Measurement involves collecting objective data using various instruments such as test and non-test (Zainal, 2020). Assessments serve not only as a measure of student achievement, but also as feedback for teachers in evaluating the effectiveness of learning (Satria, 2024). Therefore, it is important for educators to have competence in developing measurement tools that are able to measure critical thinking skills validly and reliably. Unfortunately, the reality on the ground shows that the practice of assessing mathematical critical thinking skills is still not optimal. The results of a survey of mathematics teachers in four junior high schools (SMP) in North Bahar District show that the measurement of critical thinking skills is often based only on observations during the learning process, without systematic measuring tools. The questions used in the assessment have not yet referred to the indicators of critical thinking skills, both in terms of interpretation, analysis, and evaluation.

To address this problem, it is necessary to develop a special test instrument that measures students' mathematical critical thinking skills. The description test is more appropriate than objective tests such as multiple choice because it is able to dig deep into students' thinking processes and produce rich data to analyze (Arini *et al*., 2024). In order for a test to be of quality, its validity and reliability must be tested. Furthermore, test results are analyzed using a model that provides a clear picture of students' mathematical critical thinking skills. Evaluation and improvement of question items are carried out with the right analytical approach to ensure optimal assessment instruments.

In this context, the *Item Response Theory* (IRT) is a more appropriate choice because it is able to provide a more accurate analysis of the characteristics of each question item and students' abilities. According to Sapphire *et al*., (2024) IRT allows for a more in-depth evaluation than the classical approach, as it considers the relationship between the characteristics of the question and the ability of the individual respondents. The IRT model helps improve the quality of the test by ensuring that the questions actually measure the desired ability (Falani *et al*., 2020a).

GPCM is one of the approaches in IRT used to analyze exam data with a graded score scale. Research by (Dewanti *et al*., 2021) shows that GPCM is very useful in math literacy tests that have several levels of difficulty, thus providing more representative results. According to Harsana & Lumenyela (2023) Unlike the classic model that only considers correct or incorrect answers, GPCM in IRT provides a better estimate of students' ability in question-based questions. *constructed-response* compared to classic models. GPCM can provide a more accurate picture of a person's abilities based on the responses given, so that the results of the analysis are more representative and informative (Luo, 2018).

The lack of attention in previous research on the assessment of mathematical critical thinking skills, especially in ratio materials and with the GPCM approach, has resulted in the limited availability of instruments that can accurately represent students' abilities. This condition has direct implications for the low quality of classroom learning evaluation and the lack of optimal effectiveness of pedagogical interventions, which ultimately has an impact on the overall quality of mathematics learning. Therefore, this study focuses on the assessment of mathematical critical thinking skills with the GPCM approach, as an effort to improve the accuracy of assessment and contribute to improving the quality of education.

RESEARCH METHODS

This study aims to measure students' mathematical critical thinking skills using the GPCM approach, in order to improve accuracy in the assessment process. This research was carried out from April to May 2025 at one of the junior high schools in North Bahar District, Muaro Jambi Regency. The subjects of the study are all grade VII students in the even semester of the 2024/2025 Academic Year from four State Junior High Schools in the sub-district, with a total of 116 students. This study uses a total sampling technique, where the entire population is sampled.

The research approach applied is quantitative with experimental methods. Data was collected through a written test in the form of a description, which consisted of two questions developed based on the FRISCO critical thinking skill indicator and associated with a specific mathematical material, namely Ratio. Before the analysis, the test instrument has gone through validity and reliability tests to ensure that each question item has an adequate level of accuracy and consistency in measuring the targeted ability.

The data analysis technique was carried out using the IRT model with the GPCM approach. GPCM is one of the IRT models for polythmus data. For test instruments with multiple levels of achievement, the polynomial IRT model such as GPCM A more appropriate approach (Aji & Retnawati, 2024). GPCM is used with the aim of showing the estimation of item parameters and students' abilities. The mathematical model is as follows.

$$P(X_i = x|\theta) = \frac{exp\left(\sum_{j=0}^{x} \alpha_i\left(\theta - \delta_{ij}\right)\right)}{\sum_{k=0}^{m_i} exp\left(\sum_{j=0}^{k} \alpha_i\left(\theta - \delta_{ij}\right)\right)}$$

Where:

$P(X_i = x|\theta)$ = the probability of a person with the ability to respond toθ

Categories for the following itemsxi

a_i = slope parameters (discrimination) for the item to-i

δ_{ij} = threshold parameter (difficulty level) for the category in the first item.ji

m_i = the number of categories in the first item$i$

θ = ability (*Trait*) individual (Falani *et al.*, 2020b).

As for the computational process, the software programs used in this case are SPSS 26 and PARSCALE 4.1.

RESULTS AND DISCUSSION

RESEARCH RESULTS

This research is based on IRT theory by applying GPCM, which is used to describe the difficulty level of question items and the ability of test participants. In the GPCM approach, there are three main assumptions, namely unidimensional tests, local independence, and parameter invariance that must be met.

Results of Model Prerequisite Test Analysis

The model prerequisite analysis includes three main aspects: a unidimensional test, a local independence test, and an invariance test of interrelated parameters. If the unidimensional test is met, then the local independence and parameter invariance tests are also considered to be met (Bahar & Retnawati, 2022). The unidimensional test aims to ensure that the instrument only measures one aspect, namely students' mathematical critical thinking skills. Based on Kaiser's criteria in exploratory factor analysis (EFA), an instrument is said to be unidimensional if the eigenvalue of the first component is more than 1, while the next component is less than 1. These results are typically visualized with a plot scree that shows a sharp decline after the first component, corroborating the conclusion of the unidimensionality of the data.

Table 1. Total Variance

Factor	*Initial Eigenvalues*			*Extraction Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	7,411	61,759	61,759	6,998	58,319	58,319
2	0,674	5,619	67,378			
3	0,645	5,377	72,755			
4	0,558	4,649	77,405			
5	0,502	4,183	81,588			
6	0,425	3,538	85,126			
7	0,381	3,175	88,302			
8	0,352	2,931	91,233			
9	0,300	2,500	93,733			
10	0,283	2,360	96,093			
11	0,249	2,078	98,170			
12	0,220	1,830	100,000			

Extraction Method: Principal Axis Factoring.

Based on the output of SPSS 26 Table 1, the results of *dimension reduction* for the tested data show that *principal axis factoring* extracts data into a number of factors with more than one eigenvalue. The data produced the main factor with a total *variance explained* of 61,7%, the second factor only contributed a total *variance explained* of 5,6%, while the rest had a contribution of total *variance explained* which ranged from less than 6,0%. The following Figure 1 is a graph of *the scree plot* data.

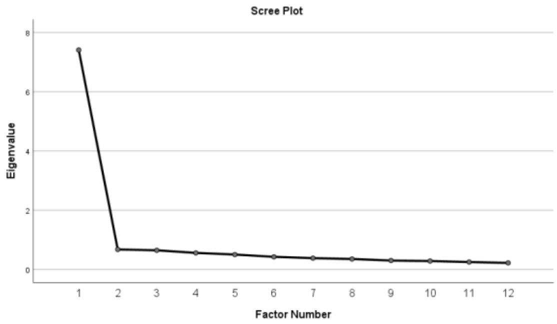

Figure 1. *Scree Plot* Data GPCM

The results of the calculation of factor analysis on the tested data and *the Scree Plot* in Figure 1 show that the main factors of each data can explain most of the total variance. So it can be concluded that the test items used are unidimensional. Most of the question items form a factor that can be called mathematical critical thinking skills.

Results of Model Fit Test Analysis (*Person Fit Order)*

Before estimating the skill parameters of the test participants, a model fit test must be carried out.

Table 2. Model Fit Test (*Person Fit Order)*

BLOCK	ITEM	CHI-SQUARE	D.F.	PROB.
BLOCK	0001	6,30831	8,	0,614
BLOCK	0002	5,68491	8,	0,684
BLOCK	0003	3,42214	9,	0,945
BLOCK	0004	7,42682	9,	0,594
BLOCK	0005	8,71319	9,	0,465
BLOCK	0006	9,81821	10,	0,457
BLOCK	0007	6,19727	10,	0,799
BLOCK	0008	1,16463	8,	0,192
BLOCK	0009	0,92243	9,	0,281
BLOCK	0010	9,95173	9,	0,018
BLOCK	0011	8,47612	8,	0,388
BLOCK	0012	3,56255	9,	0,138
TOTAL		111,64833	106,	0,335

After the unidimensional prerequisite test is carried out, then a model fit test is carried out. Table 2 of the results of the match test on the test item with the help of PARSCALE 4.1 are obtained *item fit statistics* value $\mathcal{X}^2$ GPCM is 111. (*p-value*= *0,335).* Value *p-value* (PROB.) must be above the general significance limit of 0.05 which indicates the overall fit of the model (Niken *et al*., 2024). Total value *p-value* obtained is 0,335. These findings indicate that there are no question items that deviate significantly from the model, so it can be said that all items are compatible and match the GPCM model.

Results of Estimation of Parameters of Mathematical Critical Thinking Skills (*Person Measure*)

Table 3 presents a breakdown of each student's score, which shows the range of students' mathematical critical thinking skills ranging from -2,23 to 2,35. Based on the results of *the analysis of the person measure* of 116 students, the distribution of students' abilities was divided into three categories, namely high ability (*measure* > 1.5): 7,92%, medium ability (0,5 ≤ *measure* ≤ 1,5): 28.71%, low ability (*measure* < 0,5): 63,37%. These findings indicate that the proportion of students with low ability is slightly more dominant compared to the other two categories. Table 3 presents the value *p* indicates the *person* and *m* indicates the *measure*.

Table 3. *Person Measure*

p	m	p	m	p	m	p	m	p	m	p	m
23	2,35	61	0,92	88	0,31	14	-0,13	32	-0,57	78	-1,05
92	2,35	65	0,91	38	0,29	26	-0,14	76	-0,60	70	-1,06
94	2,30	6	0,86	51	0,22	89	-0,14	69	-0,61	87	-1,09
5	2,17	40	0,85	84	0,22	63	-0,16	93	-0,61	41	-1,18
86	2,10	9	0,84	112	0,19	106	-0,17	98	-0,65	57	-1,18
72	1,59	73	0,83	8	0,13	20	-0,26	3	-0,66	48	-1,20
19	1,55	64	0,82	100	0,13	42	-0,29	77	-0,67	12	-1,34
110	1,52	90	0,82	103	0,13	97	-0,31	21	-0,69	114	-1,45
11	1,33	83	0,76	27	0,11	105	-0,34	66	-0,72	37	-1,46
75	1,32	113	0,73	49	0,10	7	-0,36	4	-0,76	116	-1,50
10	1,26	96	0,72	46	0,08	53	-0,37	104	-0,76	101	-1,58
50	1,22	99	0,71	91	0,06	44	-0,38	111	-0,80	68	-1,77
74	1,21	107	0,70	62	0,05	58	-0,42	35	-0,81	33	-2,16
55	1,19	59	0,60	13	0,02	102	-0,44	24	-0,88	34	-2,20
54	1,18	28	0,57	16	0,02	25	-0,49	31	-0,88	17	-2,23
2	1,17	22	0,52	95	0,01	18	-0,51	82	-0,91	43	-2,23
15	1,10	29	0,52	67	0,01	115	-0,52	39	-0,99		
108	1,04	36	0,44	81	-0,01	60	-0,53	71	-0,99		
30	1,02	47	0,35	80	-0,03	79	-0,54	45	-1,03		
85	1,01	52	0,33	1	-0,04	109	-0,56	56	-1,03		

Evaluation of students' skills plays a role in revealing students who have high abilities and those who show inconsistent answer patterns, such as answering carelessly, guessing, or cheating. This assessment process utilizes *Person Measure* to estimate the skill level based on the response to each question item, which is expressed in logit units. This score reflects the relative position of students on an interval scale equivalent to the difficulty scale of the problem, thus allowing for an objective comparison between students (Prayoga *et al*., 2024).

Item Measure Parameter Estimation Results

Table 4 presents the *Parscale 4.1 phase 2* outlet.

Table 4. *Item Measure*

ITEM	BLOCK	SLOPE	S.E.	LOCATION	S.E.	GUESSING	S.E.
0001	1	1,134	0,076	-0,119	0,033	0,000	0,000
0002	2	1,157	0,095	-0,102	0,031	0,000	0,000
0003	3	1,062	0,081	-0,104	0,031	0,000	0,000
0002	4	1,012	0,074	-0,029	0,033	0,000	0,000
0004	5	1,534	0,127	-0,068	0,026	0,000	0,000
0005	6	1,324	0,116	0,238	0,028	0,000	0,000
0006	7	1,153	0,091	-0,048	0,030	0,000	0,000
0008	8	0,928	0,071	-0,059	0,036	0,000	0,000
0009	9	1,079	0,066	-0,123	0,031	0,000	0,000
0010	10	1,287	0,108	-0,021	0,029	0,000	0,000
0011	11	1,307	0,112	-0,055	0,030	0,000	0,000
0012	12	0,855	0,058	0,084	0,038	0,000	0,000

Based on Table 4, the analysis of item parameters using IRT GPCM was obtained for parameter estimates for 12 question items. The estimated parameters include the discrimination parameter (*slope*/a), the difficulty parameter (*location*/b), and the guessing parameter (*guessing*/c). In general, the discriminating value (a) for all items is in the range of 0,855 to 1,534, which indicates that all grains have moderate to very high discriminating power. The difficulty value (b) ranges from -0,021 to 0,238, which indicates that these items are on a moderate difficulty. There are no very easy or very difficult items, so all items tend to be suitable for participants with average (medium) ability.

Meanwhile, all items have a guess parameter value (c) of 0,000, which indicates that the model used is 2 *Parameter Logistic Model* (2PL), or *guessing* is not taken into account in the estimate. Thus, all question items can be said to be of good quality, with a good level of discrimination and a balanced level of difficulty. This supports the assumption that the instruments used can measure participants' abilities effectively and accurately.

Test Information Function

The test information function presents data related to the characteristics of a student's ability. There is an inverse relationship between the item information function and the standard measurement error (*Standard Error Measurement*); This means that the smaller the measurement error rate, the higher the information that the test item can provide (Falani *et al*., 2020b).

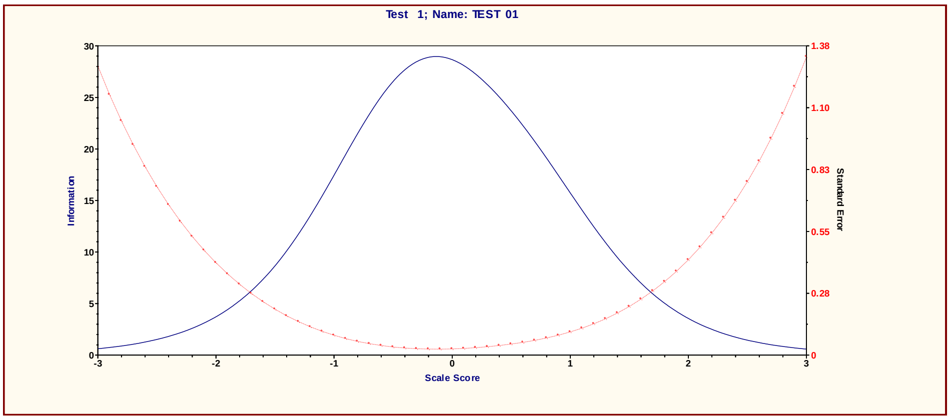

Figure 2. ***Test Information Function***

Figure 2 shows that the question items provide information optimally on the scale *score*, which is -1 to 1. The maximum information value occurs on a score scale of around 0, with an information value of around 25. This question item has the characteristics of information that is spread evenly at various levels of ability. However, the distribution of information indicating that the most informative instrument at the intermediate level is indicating the need for the development of additional items that specifically target students with low abilities.

The practical implication of these findings is that in a classroom context, teachers or instrument developers can consider the preparation of questions with lower levels of difficulty but still measure the critical thinking aspect, in order to provide greater opportunities for low-skilled students to demonstrate their potential optimally and obtain more balanced test results.

DISCUSSION

IRT is a statistical framework used to model the relationship between an individual's latent ability and their response to test items. One of the IRT models for polytomy data is GPCM. The GPCM model has a number of assumptions that need to be met first so that the results of the analysis do not produce bias. According to Liu (2024), one of the main objectives of the test of assumptions in IRT is to ensure that the data analyzed has met the basic principles on which the model is based. In this study, the three main assumptions namely unidimensionality, local independence, and parameter invariance test have been successfully met, which shows that the data is feasible to be analyzed using the GPCM model.

In addition to assumption testing, model fit testing is also an important step in IRT implementation. Through this test, it can be found whether the model structure used is in accordance with the data obtained from the field. If the model shows a good match, then parameters such as individual abilities and question characteristics can be validly and reliably estimated (Falani, 2023). This is in accordance with the main objective of the GPCM test which is to assess the suitability between the data collected and the expected model structure. Based on the results of the study, evidence was obtained that the instrument used was suitable for analysis using the GPCM model.

The results of estimating the parameters of mathematical critical thinking skills (*person measure*) in this study shows that the distribution of students' abilities that are varied and dominant is in the low category. Based on the results of the analysis of 116 students, most of the test participants were in the low ability range, while only a small number were classified as high. In the framework of item response theory, the value of this person measure is calculated based on the pattern of students' responses to question items with a certain level of difficulty and discrimination, which is expressed in a logit scale (Falani *et al*., 2020b; Luo, 2018). These scores reflect an objective estimate of students' latent abilities, where students who give consistent answers to questions with medium to high difficulty score higher, while students who answer with inconsistent patterns or show low conjecture score lower. Score range *person measure* In this study, it is shown that most students have not shown a strong mastery of the FRISCO indicator in mathematical critical thinking, thus illustrating the need for more intensive learning interventions. In addition, the diversity of scores shows that the instruments used have good discriminating power in distinguishing students' ability levels, strengthening the validity of the measurement tools used with the GPCM approach.

CONCLUSIONS AND SUGGESTIONS

This research empirically proves that the GPCM approach is effective in measuring students' mathematical critical thinking skills more accurately and representatively. The developed instrument has fulfilled three main assumptions in the IRT model, namely unidimensionality, local independence, and parameter invariance. The results of the model fit analysis showed that all question items were in accordance with the GPCM model, with moderate to high discrimination parameters and moderate difficulty levels. Student ability estimates show a fairly even distribution, with the majority of students in the low ability category. This shows that instruments have a good ability to distinguish the abilities of students from various levels. Overall, the instruments developed were declared valid and reliable to measure mathematical critical thinking skills. Further research can explore the use of digital platforms or *computer-based testing* for adaptive and efficient measurement of critical thinking skills.

REFERENCES

Adhikari, N. P. (2023). The Use of Critical Thinking for Organizational Decisions. *GS WOW: Wisdom of Worthy Research Journal*, *1*(1), 47–55. https://doi.org/https://doi.org/10.5281/zenodo.10440809

Aji, N. W., & Retnawati, H. (2024). Perbandingan Kecocokan Model Analisis Teori Respons Butir Pada Data Politomi. *Jurnal Statistika Dan Sains Data*, *1*(2), 31–47. https://doi.org/10.21831/jssd.v1i2.18484

Arini, R., Riansi, E. S., Fadillah, I., & Hasan, M. K. (2024). Peran Tes Objektif dan Subjektif Dalam Mengukur Keterampilan Berfikir Matematis Siswa Sekolah Dasar. *Jurnal Inovasi Dan Riset Pendidikan Matematika*, *5*(3), 236–244. http://www.jurnal.untirta.ac.id/index.php/wilangan%0Ap-ISSN

Atiaturrahmaniah, A., Arnyana, I. B. P., & Suastra, I. W. (2022). Peran Model Science, Technology, Engineering, Arts, And Math (STEAM) Dalam Meningkatkan Berpikir Kritis Dan Literasi Sains Siswa Sekolah Dasar. *JPGI (Jurnal Penelitian Guru Indonesia)*, *7*(4), 368–375. https://doi.org/http://dx.doi.org/10.29210/022537jpgi0005

Bahar, R., & Retnawati, H. (2022). Analisis Karakteristik Soal Kemampuan Koneksi Matematika Penskoran Politimus. *JURNAL TARBIYAH*, *29*, 195. https://doi.org/10.30829/tar.v29i2.1650

Cahayu, A., Siburian, J., Hamidah, A., & Falani, I. (2024). A Rasch Model Analysis in Measuring students' Critical Thinking Skills in science learning model Problem-Based Learning (PBL) Based on Local Wisdom Article history Abstract Keyword Rasch Model Critical Thinking Science Learning Problem-Based Local Wisdom. *Jurnal Pendidikan Sains Universitas Muhammadiyah Semarang*, *12*(1), 786–2339. https://doi.org/https://doi.org/10.26714/jps.12.1.2024.1-8

Dewanti, S. S., Yogyakarta, U. N., Sunan, U. I. N., & Yogyakarta, K. (2021). The Application of Item Response Theory in Analysis of Characteristics of Mathematical Literacy Test Items. *İlköğretim Online*, *20*(1), 1226–1237. https://doi.org/10.17051/ilkonline.2021.01.119

Falani, I. (2023). Desain dan Validasi Aplikasi Tes Literasi Matematika Berbasis Komputer dengan Pendekatan Item Response Theory. *Edu-Sains: Jurnal Pendidikan Matematika Dan Ilmu Pengetahuan Alam*, *12*(2), 16–28. https://doi.org/10.22437/jmpmipa.v12i2.26716

Falani, I., Akbar, M., & Naga, D. S. (2020a). Comparison of the Accuracy of Item Response Theory Models in Estimating Student â€™ s Ability. *Journal of Educational Science and Technology*, *6*(2), 178–184. https://doi.org/https://doi.org/10.26858/est.v6i2.13295

Falani, I., Akbar, M., & Naga, D. S. (2020b). The Precision of Students' Ability Estimation on Combinations of Item Response Theory Models. *International Journal of Instruction*, *13*(4), 545–558. https://doi.org/https://doi.org/10.29333/iji.2020.13434a

Harsana, F. N., & Lumenyela, R. A. (2024). Comparison Of Item Characteristic Analysis Models Of Reading Literacy Test With Polytomous Item Response Theory. *REID (Research and Evaluation in Education)*, *10*(2).

https://doi.org/https://doi.org/10.21831/reid.v10i2.77852 INTRODUCTION

Liu, D. T. (2024). A Scoping Review Of Rasch Analysis And Item Response Theory In Otolaryngology: Implications And Future Possibilities. In *Laryngoscope Investigative Otolaryngology* (Vol. 9, Issue 1). https://doi.org/10.1002/lio2.1208

Luo, Y. (2018). Parameter Recovery with Marginal Maximum Likelihood and Markov Chain Monte Carlo Estimation for the Generalized Partial Credit Model. *National Center for Assessment in Saudi Arabia*, 1–36. https://doi.org/https://doi.org/10.48550/arXiv.1809.07359

Mardhiyah, R. H., Aldriani, S. N. F., Chitta, F., & Zulfikar, M. R. (2021). Pentingnya Keterampilan Belajar Di Abad 21 Sebagai Tuntutan Dalam Pengembangan Sumber Daya Manusia. *Lectura: Jurnal Pendidikan*, *12*(1), 29–40. https://doi.org/http://dx.doi.org/10.31849/lectura.v12i1.5813

Muraki, E. (1992). A Generalized Partial Credit Model: Application Of An EM Algorithm Index Terms: Item Response Model, National Assess-Ment Of Educational Progress, Nominal Response Model, Partial Credit Model, Polytomous Response Model, Rating Scale Model. *Applied Psychological Measurement*, *Uj 1*, 159–176. https://doi.org/https://hdl.handle.net/11299/115645

Mursidah, M., Rosjanuardi, R., & Juandi, D. (2023). Kemampuan Berpikir Kritis Dalam Pemecahan Masalah Matematika; Systematic Literatur Review. *JPMI (Jurnal Pembelajaran Matematika Inovatif)*, *6*(4), 1421–1430. https://doi.org/https://doi.org/10.22460/jpmi.v6i4.17933

Niken, Somayasa, W., Budiman, H., Ruslan, & Adi Wibawa, G. N. (2024). Uji Goodness of Fit Dengan Statistik Pulkstenis - Robinson Dalam Model Regresi Logistik Ordinal. *Jurnal Matematika Komputasi Dan Statistika*, *4*(1), 571–576. https://doi.org/10.33772/jmks.v4i1.72

Prayoga, K. P., Suryana, D., Supriatna, M., & Budiman, N. (2024). Penggunaan Rasch Model Untuk Menganalisis Konstruk Instrumen Kontrol Diri Pada Siswa Sekolah Menengah. *G-Couns: Jurnal Bimbingan Dan Konseling*, *9*(1), 367–381. https://doi.org/10.31316/gcouns.v9i1.4459

Safitri, I., Lestarani, D., Imtikhanah, R. D. N. W., Akbarini, N. R., Sari, M. W., Fitrah, M., & Hapsan, A. (2024). *Teori Pengukuran dan Evaluasi*. CV. Ruang Tentor.

Satria, M. R. (2024). Transformasi Standar Penilaian Pendidikan Dan Revitalisasi Penilaian Pembelajaran Di Indonesia. *Jurnal Penelitian Kebijakan Pendidikan*, *17*(1), 57–66. https://doi.org/10.24832/jpkp.v17i1.930

Tohani, E., & Aulia, I. (2022). Effects Of 21st Century Learning On The Development Of Critical Thinking, Creativity, Communication, And Collaboration Skills. *Journal of Nonformal Education*, *8*(1), 46–53. https://doi.org/https://doi.org/10.15294/jne.v8i1.33334

Zainal, N. F. (2020). Pengukuran, Assessment dan Evaluasi dalam Pembelajaran Matematika. *Laplace : Jurnal Pendidikan Matematika*, *3*(1), 8–26. https://doi.org/10.31537/laplace.v3i1.310

Zalukhu, H., Naibaho, T., & Simanjuntak, R. M. (2023). Evaluasi dan Penilaian pada Pembelajaran Matematika. *Journal on Education*, *6*(01), 4781–4792. https://doi.org/https://doi.org/10.31004/joe.v6i1.3634