Phi: Jurnal Pendidikan Fisika dan Terapan
Vol. 13, No. 1 (2027) 104-117
ISSN: 2460-4348, e-ISSN: 2549-7162
https://jurnal.ar-raniry.ac.id/index.php/jurnalphi/index

Article Type: *original research*

Mapping Junior High Students' Critical Thinking Skills in Temperature and Heat with Rasch Analysis

Yasiva Dortiana Lima, Achmad Samsudin*, Duden Saepuzaman, Winny Liliawati, Lilik Hasanah
Universitas Pendidikan Indonesia, Bandung, Indonesia

Correspondence E-mail: achmadsamsudin@upi.edu

ARTICLE INFO

Article History:
Received: 26 December 2025
Revised: 03 September 2026
Accepted: 22 September 2026
Published: 05 October 2026

Keywords:
Critical thinking skills, Temperature and heat, Rasch model, Physics learning

ABSTRACT

This study examined the critical thinking skills of junior high school students on the topic of temperature and heat using the Rasch measurement approach to obtain an objective estimate of their abilities. A quantitative approach was applied, involving 53 students from a junior high school in East Nusa Tenggara Province. The research instrument was an essay test consisting of eight questions developed based on critical thinking skill indicators purpose, question at issue, assumptions, point of view, information, concepts, interpretation and inference, implications and consequences. The data obtained were subsequently analysed using the Rasch model approach with the Ministep software. The test results showed that the item reliability coefficient was 0.83, which falls within the 'good' category, while the Cronbach's alpha value of 0.60 indicated acceptable internal consistency. Item suitability analysis showed that most of the questions were in the category of fit to the Rasch model. Construct validity based on Principal Component Analysis (PCA) of the residuals showed that the instrument met the assumption of unidimensionality. In addition, the analysis of Differential Item Functioning (DIF) results showed no item bias between respondent groups. Mapping students' abilities using Wright Map and analyzing response patterns using scalograms showed variations in students' critical thinking abilities and a small number of respondents with inconsistent response patterns. These findings indicate that the developed instrument is suitable for mapping students' critical thinking skills on temperature and heat material, although further development is still needed to improve measurement accuracy at the individual level.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

1. INTRODUCTION

Education plays a fundamental role in developing human capacity by equipping individuals with the knowledge, skills and competences needed to respond to social change; in this way, it can enhance the quality of a nation's human resourcess. Improving the quality of human resources through education, from primary and secondary education to higher education, is an important element in facing the challenges and developments of the Society 5.0 era (Marfu'i et al., 2019; Voogt & Roblin, 2012). One crucial factor in facing this era is the demands of 21st-century education, which encourages schools to foster critical thinking, creativity, collaboration, and communication (4C) skills as key competencies so that students can adapt and contribute optimally to Society 5.0 (Halpern & Dunn, 2023;Astini, 2022)

One of the strategic steps taken by the Indonesian government to improve the quality of education is the implementation of various curriculum policies designed to address challenges in the field of education. One of the policies currently being implemented is Merdeka Curriculum. This curriculum emphasizes adaptive, flexible, and meaningful learning experiences tailored to the needs of students, with a primary focus on strengthening critical thinking skills and problem-solving abilities (Marfu'i et al., 2019).

Critical thinking skills are essential in physics learning because students are required to engage in analytical activities, including identifying and formulating problems, collecting and processing information, analyzing data, and drawing logical and scientific conclusions (Mabruroh & Suhandi, 2017; Hidayat et al., 2019;Bhakti et al., 2025). Physics education also contributes to students' understanding of physics and the universe (Ubaidillah et al., 2023) while providing meaningful learning experiences that strengthen their critical-thinking and problem-solving abilities (Yew et al., 2016)

However, the demand for critical-thinking skills in physics learning is not always supported by assessment instruments capable of measuring these skills accurately (Tiruneh et al., 2017). Conceptually sound, valid, and reliable instruments are therefore needed to evaluate students' cognitive achievement, scientific reasoning, and critical-thinking skills (Wulandari et al., 2025; Sartono et al., 2025). To ensure accurate measurement, educators require evidence regarding the validity, reliability, and overall quality of the assessment instruments used (Harahap & Zulhimma, 2026). Rasch modeling provides a comprehensive framework for evaluating instrument quality by examining item and person fit, reliability, and construct validity in educational assessment (Gleeson et al., 2023).

Rasch analysis is a measurement method that focuses on the suitability of empirical data with the measurement model used. This approach has a different rationale from classical statistical models because it emphasizes the quality of measurement and the accuracy of data against the model. Through Rasch analysis, the quality of items can be evaluated in depth and arranged based on their level of difficulty hierarchically (Carter et al., 2022).

Several studies have utilized the Rasch model to assess higher-order thinking skills in science education. (Sabekti & Khoirunnisa, 2018) For example, the application of the Rasch model using a partial-credit approach can be used to develop an instrument for measuring critical thinking skills. The analysis is carried out by examining the fit of the test items to the model, the level of difficulty of the items, the reliability of the instrument, and the relationship between learners' abilities and the hierarchy of item difficulty using a Wright Map. In physics education, (Dwikoranto. et al., 2025) used the Rasch approach to analyse creative thinking and problem-solving skills, focusing in particular on the reliability of the instrument, item difficulty levels, the Wright Map, and differences in ability based on gender. (Sujatmika, 2025) also used Rasch modelling to characterise pupils' critical thinking skills in science learning.

These studies indicate that the Rasch model has been widely used to evaluate the quality of instruments and to map pupils' abilities. However, analyses of the fit of individual response patterns, the functioning of items across different gender groups, and students' response patterns have not yet been the subject of a comprehensive integrated analysis. Therefore, this study combines analyses of item fit, person fit, gender-based Differential Item Functioning (DIF), Wright Maps, and scalograms. But also on the consistency of response patterns and the equivalence of item functions between male and female pupils. Consequently, this study provides a more comprehensive understanding of the characteristics of lower secondary school pupils' critical thinking skills in understanding the concepts of temperature and heat through a series of diagnostic analyses using the Rasch model.

2. METHODS

This study employed a quantitative approach involving 53 middle school students from a single school. Participants were selected through purposive sampling based on the criterion that they had previously studied the topics of temperature and heat. This criterion was used to ensure that the participants had prior learning experience relevant to the material being assessed. Because all participants were from the same school, the results of this study cannot be generalized to the broader

population of junior high school students. The instrument used was an eight-item essay test adopted directly from (Febriyana S, 2023) without any modifications.s

The data obtained in this study were analyzed using Rasch modeling with the help of Ministep software. Rasch analysis covers nine aspects, including explaining individual specifications, item criteria, and Cronbach's alpha coefficients; analysis validity consisting of Rasch model validity and construct validity; item suitability analysis based on Mean Square (MNSQ), Z-Standard (ZSTD), and Point Measure Correlation (PT Measure Corr) values; item measurement sequence analysis covering item difficulty and standard deviation; Differential Item Functioning (DIF) analysis based on probability; respondent fit analysis based on MNSQ, ZSTD, and PT Measure Corr outfit values; measurement of respondent ability (the person being measured); presentation of Wright's Map; and scalogram analysis.

The item suitability criteria in the Rasch model refer to the requirement that items are considered suitable if they meet at least one of three indicators, namely an Outfit MNSQ value between 0.5 and 1.5, an Outfit ZSTD value between −2.0 and +2.0, and a PT Measure Corr value between 0.40 and 0.85. Item suitability criteria and item difficulty index categories will be presented in detail in the analysis table (Sumintono & Widhiarso, 2014)

Tabel 1. Item-Fit Criteria

Criterion	Acceptable Range	Decision	Description
Outfit mean square	0,5<MNSQ<1,5	Accepted	Outfit mean square
Outfit Z-Standard	-2,0<ZSTD+2,0	Accepted	Outfit Z-Standard
Point measure correlation	0,4 < Pt measure corr < 0,85	Accepted	Point measure correlation

Table 2 presents the item difficulty categories determined from Rasch logit measures and their standard deviation.

Table 2. Item Difficulty Categories

Logit Value	Category
Greater than +1 SD	Very difficult
From 0.00 logit to +1 SD	Difficult
From −1 SD to 0.00 logit	Easy
Less than −1 SD	Very easy

Table 3 presents the reliability interpretation criteria derived from the reliability coefficient intervals to assess the instrument's level of measurement consistency.

Table 3. Instrument Reliability Criteria

Reliability Coefficient Interval	Interpretation
< 0.67	Weak
0.67–0.80	Adequate
0.81–0.90	Good
0.91–0.94	Very good
> 0.94	Excellent

Cronbach's alpha was used to evaluate the internal consistency of the instrument. The interpretation criteria for the Cronbach's alpha coefficient are presented in Table 4.

Table 4. Alfa Cronbach

Cronbach's Alpha Interval (α)	Category
α ≥ 0.80	Very high
0.70 ≤ α < 0.80	High
0.60 ≤ α < 0.70	Good
0.50 ≤ α < 0.60	Moderate
α < 0.50	Low

3. RESULT AND DISCUSSION

Based on the Rasch modeling analysis described above, a general mapping of instrument quality, item characteristics, and results was obtained. The mapping of these analysis results was viewed from the students' responses, where this analysis not only provided quantitative data in the form of various statistical parameters but also offered a basis for understanding how the instrument works in measuring the aspects under study. Thus, this discussion will focus on an in-depth review of the main findings, particularly regarding the validity model, measurement reliability, and reinforcement of measurement results related to instrument quality and respondent ability. This discussion will be conducted by linking the research results using Rasch modeling theory and findings from previous studies to obtain a more comprehensive understanding. Figure 3.1 below maps the reliability results of measure person and measure item.

3.1 Item Reliability and Cronbach's Alpha

Figure 3 presents the results of the reliability analysis of the critical thinking skills instrument using the Rasch model, including the Cronbach's alpha (KR-20) value and item reliability. Cronbach's alpha indicates the internal consistency of the instrument by showing the extent to which the test items measure the intended construct of critical thinking skills. Meanwhile, item reliability in the Rasch analysis reflects the consistency of item performance in measuring pupils' critical thinking abilities on the measurement scale (Taber, 2018).

```
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .60  SEM = 1.33
STANDARDIZED (50 ITEM) RELIABILITY = .92

    SUMMARY OF 8 MEASURED Item
+-----------------------------------------------------------------------------+
|          TOTAL                          MODEL        INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE     S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     163.6      53.0         .00      .31       .99   -.17    .96   -.31 |
| SEM        3.2        .0         .31      .00       .16    .76    .17    .69 |
| P.SD       8.4        .0         .83      .01       .42   2.00    .46   1.82 |
| S.SD       8.9        .0         .89      .01       .45   2.14    .49   1.94 |
| MAX.     172.0      53.0        1.45      .32      1.65   2.89   1.82   2.85 |
| MIN.     149.0      53.0        -.83      .30       .47  -2.96    .43  -2.80 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .34 TRUE SD     .76  SEPARATION  2.24  Item   RELIABILITY  .83 |
|MODEL RMSE    .31 TRUE SD     .77  SEPARATION  2.47  Item   RELIABILITY  .86 |
| S.E. OF Item MEAN = .31                                                     |
+-----------------------------------------------------------------------------+
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 1. Cronbach Alpha Values and Item Reliability of the Instrument

Based on the analysis results, the Cronbach's alpha (KR-20) value of the critical thinking skills instrument is 0.60, while the item reliability value is 0.83.

A Cronbach's alpha (KR-20) value of 0.60 indicates that the internal consistency of the critical thinking skills instrument is moderate and still acceptable in exploratory research with a limited number of items. Meanwhile, an item reliability value of 0.83 indicates that the items have good stability in measuring the construct of critical thinking skills (Sumintono & Widhiarso, 2014).

3.2 Validity (Rasch Model)

Figure 2 shows the results of the item validity test using Rasch analysis, taking into account the Infit MNSQ, Outfit MNSQ and ZSTD values to assess the fit of the items to the measurement model.

SUMMARY OF 8 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	163.6	53.0	.00	.31	.99	-.17	.96	-.31
SEM	3.2	.0	.31	.00	.16	.76	.17	.69
P.SD	8.4	.0	.83	.01	.42	2.00	.46	1.82
S.SD	8.9	.0	.89	.01	.45	2.14	.49	1.94
MAX.	172.0	53.0	1.45	.32	1.65	2.89	1.82	2.85
MIN.	149.0	53.0	-.83	.30	.47	-2.96	.43	-2.80

REAL RMSE	.34	TRUE SD	.76	SEPARATION	2.24	Item	RELIABILITY	.83
MODEL RMSE	.31	TRUE SD	.77	SEPARATION	2.47	Item	RELIABILITY	.86
S.E. OF Item MEAN = .31								

Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 2. Item Validity Test Results Using the Rasch Model

Based on the analysis results, the Mean Square infit value (MNSQ) is 0.99 and the outfit MNSQ is 0.96. Meanwhile, the Z-Standard infit value (ZSTD) is −0.17 and the outfit ZSTD is −0.31.

The infit and outfit MNSQ values, which are in the range of 0.5–1.5, and the ZSTD values, which are in the range of −2 to +2, indicate that the critical thinking skill items are well suited to the Rasch model. According to (Sumintono & Widhiarso, 2014), this condition indicates that the items are able to measure the same construct consistently and in line with the Rasch model assumptions, so that the items can be used for further analysis.

3.3 Construct Validity

Figure 3 presents the PCA residual analysis of the Rasch model to examine the construct validity of the critical thinking skills instrument. The purpose of this analysis is to evaluate the extent to which the measurement tool can capture a single main construct (unidimensionality) by looking at the percentage of variance explained by the measurement and the unexplained variance in each residual contrast. The measurement tool can be considered to meet the unidimensionality assumption if the percentage of raw variance explained by the measure reaches the recommended minimum threshold and the unexplained variance in the first contrast is within an acceptable range, indicating that there are no other dimensions that are more dominant outside the main construct being measured (Sumintono & Widhiarso, 2014)).

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed	Expected
Total raw variance in observations =	11.9855	100.0%	100.0%
Raw variance explained by measures =	3.9855	33.3%	33.0%
Raw variance explained by persons =	3.2265	26.9%	26.7%
Raw Variance explained by items =	.7589	6.3%	6.3%
Raw unexplained variance (total) =	8.0000	66.7% 100.0%	67.0%
Unexplned variance in 1st contrast =	1.9139	16.0%	23.9%
Unexplned variance in 2nd contrast =	1.6099	13.4%	20.1%
Unexplned variance in 3rd contrast =	1.2622	10.5%	15.8%
Unexplned variance in 4th contrast =	1.1763	9.8%	14.7%
Unexplned variance in 5th contrast =	.8929	7.4%	11.2%

Figure 3. Construct validity of critical thinking skills

Based on Table 3, the results of PCA residual analysis in the Rasch model show that the raw variance explained by the student critical thinking skills measurement instrument is 33.3%, while the unexplained variance in the first contracts is 16.0%, with the eigenvalue in the first contrast below 2.0.

A raw variance value explained by a measure of >20% indicates that the construct of students' critical thinking skills is able to explain an adequate proportion of variance in the measurement. Although the unexplained variance value in the first contrast slightly exceeds the ideal limit of <15%, the eigenvalue value, which is still below 2.0, indicates that there is no dominant second dimension. This indicates that the instrument still measures one main construct, namely students' critical thinking skills, in accordance with the assumption of unidimensionality in the Rasch

model (Boone et al., 2014). However, the presence of residual variation indicates that the development of a student critical thinking skills instrument is still needed to improve measurement accuracy.

3.4 Misfit Item Order

The table below explains the suitability of items in the Rasch model evaluated using the Outfit Mean Square (MNSQ), Z-Standard (ZSTD), and Point Measure Correlation (PT Measure Corr) values. According to Rasch modeling criteria, an item is considered a good fit if the Outfit MNSQ value is in the range of 0.5–1.5 (ideal value 1.0), the Outfit ZSTD value is in the range of −2.0 to +2.0 (ideal value 0.0), and the PT Measure Corr value is in the range of 0.40–0.85 However, items with MNSQ and ZSTD values outside the tolerance limits are still acceptable as long as the Point Measure Correlation value is positive, as this indicates that the item still measures the same construct (Boone et al., 2014).

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA-AL CORR.	EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
5	172	53	-.83	.31	1.65	2.89	1.82	2.85	A .61	.51	56.6	73.6	S5
7	170	53	-.64	.32	1.65	2.77	1.50	1.86	B .34	.51	64.2	74.2	S7
3	149	53	1.45	.30	1.13	.64	1.11	.47	C .28	.53	75.5	74.6	S3
2	168	53	-.44	.32	.95	-.15	.92	-.25	D .64	.50	71.7	75.2	S2
4	150	53	1.36	.30	.73	-1.26	.67	-1.29	d .59	.52	83.0	75.0	S4
8	166	53	-.23	.32	.69	-1.51	.67	-1.38	c .55	.50	81.1	76.1	S8
6	168	53	-.44	.32	.65	-1.81	.57	-1.98	b .65	.50	79.2	75.2	S6
1	166	53	-.23	.32	.47	-2.96	.43	-2.80	a .49	.50	88.7	76.1	S1
MEAN	163.6	53.0	.00	.31	.99	-.17	.96	-.31			75.0	75.0	
P.SD	8.4	.0	.83	.01	.42	2.00	.46	1.82			9.8	.8	

Figure 4. Figure 6. Rasch Person Fit Statistics in Misfit Order

Based on Table 4, the results of the analysis of the suitability of the critical thinking skills instrument items show that item number 5 has an Outfit Mean Square (MNSQ) value of 1.82 and an Outfit Z-Standard (ZSTD) value of 2.85, while the Point Measure Correlation (PT Measure Corr) value is 0.61. Meanwhile, items 7, 3, 2, 4, 8, 6, and 1 have Outfit MNSQ and Outfit ZSTD values within the specified criteria range, as well as positive PT Measure Corr values.

Based on the above results, item number 5 shows Outfit Mean Square (MNSQ) and Outfit Z-Standard (ZSTD) values that are outside the Rasch model suitability criteria range, indicating a mismatch with the model. However, the positive Point Measure Correlation (PT Measure Corr) value indicates that the item still measures students' critical thinking skills in line with the main construct. This condition shows that even though the students' response patterns to this item are more varied, item number 5 still requires critical thinking skills, such as disseminating information or disseminating reasons, so it is still relevant for measuring students' critical thinking skills. Therefore, this item can still be retained with the note that editorial improvements are made so that the critical thinking cognitive demands being measured become clearer (Planinic et al., 2019) explains that items with complex contexts often show high outfit values, because students use a variety of reasoning approaches, so that the phenomenon of slight misfit on critical thinking items is reasonable and methodologically acceptable.

Meanwhile, items 7, 3, 2, 4, 8, 6, and 1 show good fit values with the Rasch model. This indicates that students' response patterns to these items are consistent with the expected level of difficulty, so that each item is able to distinguish students' critical thinking skill levels more reliably. Items that meet these suitability criteria effectively represent critical thinking skill indicators, such as the ability to interpret concepts, analyze cause-and-effect relationships, and draw logical conclusions, thereby supporting the accurate and consistent measurement of students' critical thinking skills.

3.5 Item Statistics: Measure Order

Figure 5. shows the order of item difficulty based on the item size (logit) value resulting from Rasch modeling, which is used to describe the hierarchy of item difficulty in measuring students' critical thinking skills.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	149	53	1.45	.30	1.13	.64	1.11	.47	.28	.53	75.5	74.6	A3
4	150	53	1.36	.30	.73	-1.20	.67	-1.29	.59	.52	83.0	75.0	A4
1	166	53	-.23	.32	.47	-2.96	.43	-2.80	.49	.50	88.7	76.1	A1
8	166	53	-.23	.32	.69	-1.51	.67	-1.38	.55	.50	81.1	76.1	A8
2	168	53	-.44	.32	.95	-.15	.92	-.25	.64	.50	71.7	75.2	A2
6	168	53	-.44	.32	.65	-1.81	.57	-1.98	.65	.50	79.2	75.2	A6
7	170	53	-.64	.32	1.65	2.77	1.50	1.86	.34	.51	64.2	74.2	A7
5	172	53	-.83	.31	1.65	2.89	1.82	2.85	.61	.51	56.6	73.6	A5
MEAN	163.6	53.0	.00	.31	.99	-.17	.96	-.31			75.0	75.0	
P.SD	8.4	.0	.83	.01	.42	2.00	.46	1.82			9.8	.8	

Figure 5. Item Difficulty Measures Based on the Rasch Model

Based on the above, the difficulty level of the critical thinking skills instrument items, expressed in logit units, ranges from −0.83 to 1.45 logit. The average difficulty level of the items is 0.00 logit, with a standard deviation of 0.83. The item with the highest logit value is item number 3 at 1.45 logit, while the item with the lowest logit value is −0.83 logit.

The range of item difficulty values between −2 and +2 logit indicates that the critical thinking skill items were developed at a level of difficulty consistent with the Rasch model assumption. The average item value of 0.00 logit indicates that the measurement scale is satisfactorily balanced. In addition, the standard deviation value of 0.83 indicates that there is adequate variation in item difficulty, so that the instrument is not homogeneous. Items with the highest logit values require a higher level of critical thinking skills from students, such as the ability to disseminate and spread concepts, while items with the lowest logit values require more basic critical thinking skills. This variation in difficulty levels supports the instrument's ability to measure students' critical thinking skills at various ability levels.

3.6 Differential Item Functioning (DIF)

Differential Item Functioning (DIF) analysis was conducted to determine whether the instrument items functioned differently between male and female respondents with comparable levels of critical-thinking ability. In the Winsteps specification DIF=$S3W1, S3 indicates that the grouping variable was extracted from the third character of each person label, whereas W1 indicates that the group code consisted of one character. The code L represented male respondents, while P represented female respondents. The analysis included 21 male respondents and 32 female respondents. Table 6 presents the results of the DIF analysis conducted using the Rasch model.

Table 6. Results of the Differential Item Functioning Analysis by Gender

Person Classes	Summary Dif Chi Squared	D.F	PROB	Beetwen-Class/Group Item		
				UNWTD MNSQ	ZSTD	Number Name
2	.0874	1	0.7675	.0852	-.72	1 A1
2	.0218	1	.8825	.0204	-1.07	2 A2
2	1.0636	1	.3024	1.1006	.54	3 A3
2	.5987	1	.4391	.6159	.15	4 A4
2	.2289	1	.6323	.2288	-.35	5 A5
2	.6214	1	.4305	.6323	.17	6 A6
2	.0000	1	1.0000	.0003	-1.52	7 A7
2	.0874	1	.7675	.0852	-0.72	8 A8

Note: DIF class/group specification is DIF=$S3W1.

Based on the results of the Differential Item Functioning (DIF) analysis on the critical thinking skills instrument, all items had a probability value (p-value) above 0.05. The DIF probability values for each item ranged from 0.3024 to 1.0000, as shown in Table 6. Thus, no significant differences in item functioning were found between respondent groups. According to (Linacre, 2010) an item is considered biased if the DIF probability value is less than 0.05.

The analysis results show that all items in the critical thinking skills instrument have probability values above this threshold, so there is no indication of bias between respondent groups. This shows that each item provides equal opportunities for all students to demonstrate their critical thinking skills, without favoring or disadvantaging any particular group. Thus, these items can be considered fair and suitable for measuring students' critical thinking (Boone et al., 2014; Bond & Fox, 2015).

3.7 Fit Person

Figure 6. shows the results of person fit analysis based on Rasch modeling to identify the consistency of student response patterns to the instruments used (Boone et al., 2014).

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	PTMEA EXP.	SUR-AL EXACT MATCH OBS%	SUR-AL EXACT MATCH EXP%	Person
3	24	8	1.77	.89	5.44	3.26	5.10	3.02	A .63	.30	37.5	83.1	03P
2	24	8	1.77	.89	4.31	2.73	4.70	2.83	B .40	.30	25.0	83.1	02P
18	26	8	3.18	.78	2.74	2.96	3.22	2.82	C-.59	.31	50.0	70.9	18L
17	25	8	2.53	.84	2.34	1.80	2.34	1.73	D-.29	.28	62.5	79.4	17P
4	20	8	-.66	.66	1.95	1.74	2.09	1.94	E .03	.45	25.0	65.0	04L
5	19	8	-1.07	.63	1.72	1.52	1.79	1.64	F .54	.45	12.5	56.6	05L
1	28	8	4.30	.75	1.36	1.14	1.53	1.40	G .86	.38	75.0	65.5	01L
6	24	8	1.77	.89	1.14	.43	1.01	.27	H .69	.30	75.0	83.1	06L
7	24	8	1.77	.89	1.14	.43	1.01	.27	I .69	.30	75.0	83.1	07P
11	24	8	1.77	.89	1.14	.43	1.01	.27	J .69	.30	75.0	83.1	11P
12	24	8	1.77	.89	1.14	.43	1.01	.27	K .69	.30	75.0	83.1	12L
32	23	8	1.00	.85	.99	.22	1.14	.44	L-.36	.36	87.5	80.0	33L
42	23	8	1.00	.85	.99	.22	1.14	.44	M-.36	.36	87.5	80.0	42L
46	23	8	1.00	.85	.99	.22	1.14	.44	N-.36	.36	87.5	80.0	46L
48	23	8	1.00	.85	.95	.16	1.09	.38	O-.29	.36	87.5	80.0	48P
33	22	8	.35	.77	.90	.04	1.03	.27	P .18	.41	75.0	74.4	33P
43	22	8	.35	.77	.90	.04	1.03	.27	Q .18	.41	75.0	74.4	43P
44	27	8	3.75	.74	.99	.08	.89	-.11	R .28	.34	37.5	62.5	44P
47	27	8	3.75	.74	.99	.08	.89	-.11	S .28	.34	37.5	62.5	47P
19	22	8	.35	.77	.87	-.02	.93	.12	T .29	.41	75.0	74.4	19L
22	22	8	.35	.77	.87	-.02	.93	.12	U .29	.41	75.0	74.4	22L
37	26	8	3.18	.78	.85	-.13	.80	-.21	V .23	.31	75.0	70.9	37P
38	26	8	3.18	.78	.85	-.13	.80	-.21	W .23	.31	75.0	70.9	38P
35	27	8	3.75	.74	.88	-.31	.79	-.38	X .41	.34	62.5	62.5	35P
30	26	8	3.18	.78	.78	-.43	.68	-.48	Y .37	.31	75.0	70.9	30L
31	26	8	3.18	.78	.78	-.43	.68	-.48	Z .37	.31	75.0	70.9	31P
40	26	8	3.18	.78	.78	-.43	.68	-.48	.37	.31	75.0	70.9	40L
41	26	8	3.18	.78	.78	-.43	.68	-.48	z .37	.31	75.0	70.9	41P
13	27	8	3.75	.74	.77	-.72	.69	-.67	y .53	.34	62.5	62.5	13P
15	27	8	3.75	.74	.77	-.72	.69	-.67	x .53	.34	62.5	62.5	15P
50	27	8	3.75	.74	.77	-.72	.69	-.66	w .53	.34	87.5	62.5	50P
52	28	8	4.30	.75	.73	-.87	.68	-.89	v .64	.38	75.0	65.5	52P
14	26	8	3.18	.78	.72	-.60	.63	-.61	u .44	.31	75.0	70.9	14P
16	26	8	3.18	.78	.72	-.60	.63	-.61	t .44	.31	75.0	70.9	16L
36	29	8	4.89	.79	.66	-.88	.66	-.84	s .74	.40	87.5	72.5	36P
21	25	8	2.53	.84	.62	-.51	.59	-.53	r .20	.28	87.5	79.4	21P
27	25	8	2.53	.84	.62	-.51	.59	-.53	q .20	.28	87.5	79.4	27P
29	25	8	2.53	.84	.57	-.63	.51	-.71	p .29	.28	87.5	79.4	29P
39	25	8	2.53	.84	.57	-.63	.51	-.71	o .29	.28	87.5	79.4	39P
49	25	8	2.53	.84	.57	-.63	.51	-.71	n .29	.28	87.5	79.4	49P
8	25	8	2.53	.84	.51	-.78	.44	-.90	m .38	.28	87.5	79.4	08L
9	25	8	2.53	.84	.51	-.78	.44	-.90	l .38	.28	87.5	79.4	09P
10	25	8	2.53	.84	.51	-.78	.44	-.90	k .38	.28	87.5	79.4	10L
34	30	8	5.58	.88	.44	-1.30	.34	-1.26	j .97	.39	100.0	76.9	34P
20	23	8	1.00	.85	.40	-.88	.25	-1.09	i .62	.36	87.5	80.0	20L
23	23	8	1.00	.85	.40	-.88	.25	-1.09	h .62	.36	87.5	80.0	23L
24	23	8	1.00	.85	.40	-.88	.25	-1.09	g .62	.36	87.5	80.0	24P
26	23	8	1.00	.85	.40	-.88	.25	-1.09	f .62	.36	87.5	80.0	26P
28	23	8	1.00	.85	.40	-.88	.25	-1.09	e .62	.36	87.5	80.0	28P
51	23	8	1.00	.85	.36	-.98	.22	-1.18	d .66	.36	87.5	80.0	51L
25	24	8	1.77	.89	.10	-1.87	.08	-1.88	c .00	.30	100.0	83.1	25P
45	24	8	1.77	.89	.10	-1.87	.08	-1.88	b .00	.30	100.0	83.1	45L
53	24	8	1.77	.89	.10	-1.87	.08	-1.88	a .00	.30	100.0	83.1	53L

Figure 6 Person Fit Statistics Based on the Rasch Model Misfit Order

Based on the respondent fit analysis in the Rasch model, it was found that 7 out of 53 students (13.2%) had an Outfit Mean Square (MNSQ) value greater than 1.5. These respondents were students with codes 03P, 02P, 18L, 17P, 04L, 05L, and 01L. This finding indicates that there were respondents whose answer patterns did not match the level of difficulty of the items in the critical thinking skills instrument.

Outfit Mean Square (MNSQ) values exceeding the 1.5 limit in a number of respondents indicate inconsistent response patterns to the hierarchy of difficulty levels of critical thinking skill

items. In Rasch modeling, an Outfit MNSQ value > 1.5 indicates that individual responses contain noise or deviations from model expectations, which may be caused by non-cognitive factors such as lack of accuracy, errors in understanding the questions, or unstable answering strategies (Linacre, 2010)

However, the proportion of non-compliant respondents is relatively small compared to the total number of students. (Bond & Fox, 2015; Sumintono & Widhiarso, 2014). The existence of a small number of non-compliant respondents is a normal condition in Rasch analysis and does not necessarily indicate a weakness in the instrument. Therefore, the non-conforming responses found reflect more the variation in individual student behavior in answering questions, rather than the inability of the instrument to measure the construct. Thus, this instrument can still be used to measure students' critical thinking skills as a whole (Bond & Fox, 2015; Sumintono & Widhiarso, 2014)

3.8 Ability

Figure 7. presents the results of measuring students' critical thinking skills, which were analyzed using the Rasch model. Students' abilities are expressed in logit units and ranked from highest to lowest to illustrate the distribution of students' abilities in answering critical thinking skill items. In Rasch modeling, ability (person ability) is represented as a linear measure in logit units, so that differences in ability between students can be compared objectively and independently of raw scores (Tesio et al., 2024).

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
34	30	8	5.58	.88	.44	-1.30	.34	-1.26	.97	.39	100.0	76.9	34P
36	29	8	4.89	.79	.66	-.88	.66	-.84	.74	.40	87.5	72.5	36P
1	28	8	4.30	.75	1.36	1.14	1.53	1.40	.86	.38	75.0	65.5	01L
52	28	8	4.30	.75	.73	-.87	.68	-.89	.64	.38	75.0	65.5	52P
13	27	8	3.75	.74	.77	-.72	.69	-.67	.53	.34	62.5	62.5	13P
15	27	8	3.75	.74	.77	-.72	.69	-.67	.53	.34	62.5	62.5	15P
35	27	8	3.75	.74	.88	-.31	.79	-.38	.41	.34	62.5	62.5	35P
44	27	8	3.75	.74	.99	.08	.89	-.11	.28	.34	37.5	62.5	44P
47	27	8	3.75	.74	.99	.08	.89	-.11	.28	.34	37.5	62.5	47P
50	27	8	3.75	.74	.77	-.72	.69	-.66	.53	.34	87.5	62.5	50P
14	26	8	3.18	.78	.72	-.60	.63	-.61	.44	.31	75.0	70.9	14P
16	26	8	3.18	.78	.72	-.60	.63	-.61	.44	.31	75.0	70.9	16L
18	26	8	3.18	.78	2.74	2.96	3.22	2.82	-.59	.31	50.0	70.9	18L
30	26	8	3.18	.78	.78	-.43	.68	-.48	.37	.31	75.0	70.9	30L
31	26	8	3.18	.78	.78	-.43	.68	-.48	.37	.31	75.0	70.9	31P
37	26	8	3.18	.78	.89	-.13	.80	-.21	.23	.31	75.0	70.9	37P
38	26	8	3.18	.78	.89	-.13	.80	-.21	.23	.31	75.0	70.9	38P
40	26	8	3.18	.78	.78	-.43	.68	-.48	.37	.31	75.0	70.9	40L
41	26	8	3.18	.78	.78	-.43	.68	-.48	.37	.31	75.0	70.9	41P
8	25	8	2.53	.84	.51	-.78	.44	-.90	.38	.28	87.5	79.3	08L
9	25	8	2.53	.84	.51	-.78	.44	-.90	.38	.28	87.5	79.3	09P
10	25	8	2.53	.84	.51	-.78	.44	-.90	.38	.28	87.5	79.3	10L
17	25	8	2.53	.84	2.34	1.80	2.34	1.73	-.29	.28	62.5	79.3	17P
21	25	8	2.53	.84	.62	-.51	.59	-.53	.20	.28	87.5	79.3	21P
27	25	8	2.53	.84	.62	-.51	.59	-.53	.20	.28	87.5	79.3	27P
29	25	8	2.53	.84	.57	-.63	.51	-.71	.29	.28	87.5	79.3	29P
39	25	8	2.53	.84	.57	-.63	.51	-.71	.29	.28	87.5	79.3	39P
49	25	8	2.53	.84	.57	-.63	.51	-.71	.29	.28	87.5	79.3	49P
2	24	8	1.77	.89	4.31	2.73	4.69	2.83	.40	.30	25.0	83.1	02P
3	24	8	1.77	.89	5.44	3.26	5.10	3.02	.63	.30	37.5	83.1	03P
6	24	8	1.77	.89	1.14	.43	1.01	.27	.69	.30	75.0	83.1	06L
7	24	8	1.77	.89	1.14	.43	1.01	.27	.69	.30	75.0	83.1	07P

11	24	8	1.77	.89	1.14	.43	1.01	.27	.69	.30	75.0	83.1	11P
12	24	8	1.77	.89	1.14	.43	1.01	.27	.69	.30	75.0	83.1	12L
25	24	8	1.77	.89	.10	-1.87	.08	-1.88	.00	.30	100.0	83.1	25P
45	24	8	1.77	.89	.10	-1.87	.08	-1.88	.00	.30	100.0	83.1	45L
53	24	8	1.77	.89	.10	-1.87	.08	-1.88	.00	.30	100.0	83.1	53L
20	23	8	1.00	.85	.40	-.88	.25	-1.09	.62	.36	87.5	80.0	20L
23	23	8	1.00	.85	.40	-.88	.25	-1.09	.62	.36	87.5	80.0	23L
24	23	8	1.00	.85	.40	-.88	.25	-1.09	.62	.36	87.5	80.0	24P
26	23	8	1.00	.85	.40	-.88	.25	-1.09	.62	.36	87.5	80.0	26P
28	23	8	1.00	.85	.40	-.88	.25	-1.09	.62	.36	87.5	80.0	28P
32	23	8	1.00	.85	.99	.22	1.14	.44	-.38	.36	87.5	80.0	33L
42	23	8	1.00	.85	.99	.22	1.14	.44	-.38	.36	87.5	80.0	42L
46	23	8	1.00	.85	.99	.22	1.14	.44	-.38	.36	87.5	80.0	46L
48	23	8	1.00	.85	.95	.16	1.09	.38	-.29	.36	87.5	80.0	48P
51	23	8	1.00	.85	.36	-.98	.22	-1.18	.66	.36	87.5	80.0	51L
19	22	8	.35	.77	.87	-.02	.93	.12	.25	.41	75.0	74.4	19L
22	22	8	.35	.77	.87	-.02	.93	.12	.25	.41	75.0	74.4	22L
33	22	8	.35	.77	.90	.04	1.03	.27	.18	.41	75.0	74.4	33P
43	22	8	.35	.77	.90	.04	1.03	.27	.18	.41	75.0	74.4	43P
4	20	8	-.66	.66	1.95	1.74	2.09	1.94	.03	.45	25.0	65.0	04L
5	19	8	-1.07	.63	1.72	1.52	1.79	1.64	.54	.45	12.5	56.6	05L
MEAN	24.7	8.0	2.24	.81	.99	-.13	.96	-.16			75.0	75.0	
P.SD	2.1	.0	1.37	.06	.92	1.08	.96	1.11			19.4	7.2	

Figure 7. Order of Respondents' Abilities Based on Measure Values (Person Measure Order) Rasch Model)

Based on figure 7. students' ability scores on the critical thinking skills instrument, expressed in logit units, ranged from −1.07 to 5.58 logit. Students with the highest ability scores had a value of 5.58 logit, while students with the lowest scores had a value of −1.07 logit. These findings indicate that there is variation in students' abilities in answering critical thinking skill items.

The relatively wide range of student ability scores indicates that the instrument is able to clearly distinguish the levels of students' critical thinking skills. Students with high logit scores demonstrate better critical thinking skills, such as the ability to disseminate and provide logical reasons for the problems given. Conversely, students with low logit scores demonstrate limited critical thinking skills, especially in understanding and solving problems systematically. This variation in skill levels reflects differences in characteristics and mastery of critical thinking skills among students, not weaknesses in the measurement instrument.

3.9 Wright Map

Figure 8. Displaying the Wright map (Person–Item Map) resulting from Rasch model analysis, which maps student abilities and the difficulty levels of critical thinking skill items on the subject of temperature and heat in the same logit units. The distribution of student abilities is displayed on the left side of the map, while the hierarchy of item difficulty levels is displayed on the right side, covering the categories of easy, medium, and difficult. This map provides information about the suitability between student abilities and question characteristics, and shows the extent to which the instrument is able to comprehensively measure variations in student abilities. Person–item mapping in the Rasch model allows for the evaluation of the relationship between student abilities and the difficulty level of questions on the same measurement scale, so that the quality of the instrument and the distribution of student abilities can be analyzed objectively and meaningfully (Ibrahim, 2025)

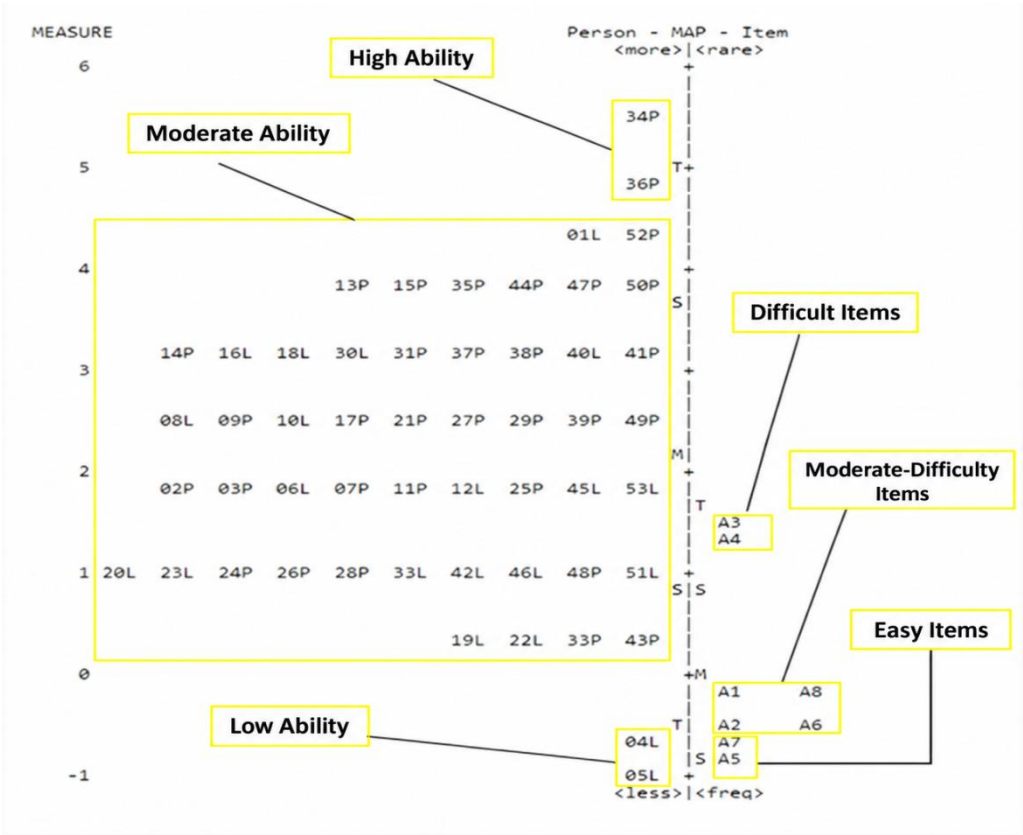


Figure 8. Wright map

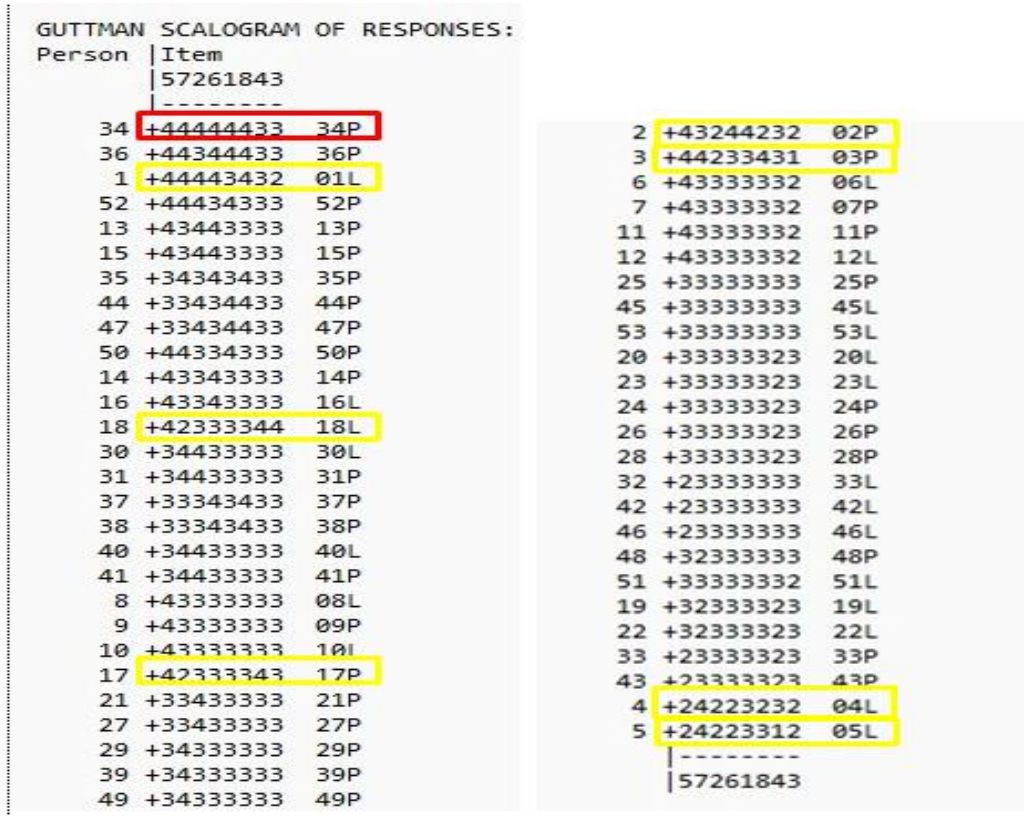
The Wright Map (Person-Item Map) illustrates the distribution of pupils' abilities and the hierarchy of difficulty levels for critical thinking questions on the same logit scale. The mapping results show that the majority of pupils have abilities above 0 logit. The highest ability levels are occupied by pupils with codes 34P and 36P, who are at the highest logit level. Students with moderate abilities are spread across a range of approximately 0 to 4 logits, while some students are at ability positions below 0 logits. In terms of items, items A3 and A4 are at a higher level of difficulty, while items A1, A2, A5, A6, and A7 are at a low to moderate level of difficulty.

Wright's map shows the relationship between student ability and the level of difficulty of critical thinking questions on the same logit scale. The distribution of student abilities, which is predominantly above 0 logit, indicates that most students have relatively higher critical thinking skills compared to the level of difficulty of the questions given. Students with high abilities are able to answer items with high levels of difficulty, which require critical thinking processes such as analysis, evaluation, and advanced conclusions. Meanwhile, students with moderate abilities tend to be able to work on items with low to moderate levels of difficulty, which represent critical thinking skills at a basic to intermediate level. The existence of students with low abilities who are not yet able to complete some items shows the variation in critical thinking abilities among students. This variation reflects differences in students' characteristics and mastery of critical thinking skills, and provides an indication that the instrument is able to adequately distinguish between students' levels of critical thinking ability.

3.10 Scalogram

Figure 9. Presenting the Guttman Scalogram of Responses from the analysis using the Rasch model. This scalogram displays the response patterns of each student to critical thinking skill items that have been arranged based on the order of item difficulty, thus forming a hierarchical response pattern. The presentation of the scalogram aims to provide an initial overview of the consistency of students' answer patterns to the question difficulty hierarchy by comparing the observed response patterns with the expected response patterns based on the differences between the respondents'

abilities and the difficulty levels of the items. In the Rasch model, examining these hierarchical response patterns serves as an initial diagnosis of data quality before conducting further analyses, such as fit analysis and student ability mapping (Tennant et al., 2004).



```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |57261843
       |--------
    34 +44444433  34P
    36 +44344433  36P
     1 +44443432  01L
    52 +44434333  52P
    13 +43443333  13P
    15 +43443333  15P
    35 +34343433  35P
    44 +33434433  44P
    47 +33434433  47P
    50 +44334333  50P
    14 +43343333  14P
    16 +43343333  16L
    18 +42333344  18L
    30 +34433333  30L
    31 +34433333  31P
    37 +33343433  37P
    38 +33343433  38P
    40 +34433333  40L
    41 +34433333  41P
     8 +43333333  08L
     9 +43333333  09P
    10 +43333333  10L
    17 +42333343  17P
    21 +33433333  21P
    27 +33433333  27P
    29 +34333333  29P
    39 +34333333  39P
    49 +34333333  49P
     2 +43244232  02P
     3 +44233431  03P
     6 +43333332  06L
     7 +43333332  07P
    11 +43333332  11P
    12 +43333332  12L
    25 +33333333  25P
    45 +33333333  45L
    53 +33333333  53L
    20 +33333323  20L
    23 +33333323  23L
    24 +33333323  24P
    26 +33333323  26P
    28 +33333323  28P
    32 +23333333  33L
    42 +23333333  42L
    46 +23333333  46L
    48 +32333333  48P
    51 +33333332  51L
    19 +32333323  19L
    22 +32333323  22L
    33 +23333323  33P
    43 +23333323  43P
     4 +24223232  04L
     5 +24223312  05L
       |--------
       |57261843
```

Figure 9. Scalogram

Based on Figur 9. the Guttman response scalogram is displayed, which maps the students' response patterns to the items based on the order of item difficulty. This scalogram shows the order of each student's answers from the easiest to the most difficult items, thus enabling consistency in response patterns. The analysis results show that student number 34P has a consistent response pattern, as indicated by the response sequence 44444433. This pattern reflects that the student is able to answer the items gradually from low to high difficulty levels in accordance with the item hierarchy. In addition, seven students were found to have anomalous response patterns, namely students with respondent numbers 01L, 18L, 17P, 02P, 03P, 04L, and 05L. The response patterns of these students did not fully follow the order of item difficulty. An example of an anomalous response pattern is shown by a response sequence such as 44443432, which indicates irregularity in answering the questions.

Consistent response patterns, as demonstrated by student 34P, indicate that the student has abilities that are in line with the difficulty level of the items and is able to maintain consistency in responses in accordance with cumulative estimates on the Guttman scale. This reflects the student's understanding of the material and the stability of their strategy in answering questions. Conversely, the anomalous response patterns shown by some students indicate a mismatch between the students' abilities and their responses to the difficulty level of the items. In the context of the Rasch model, non-sequential responses may reflect several possibilities, such as conceptual misunderstanding, carelessness in answering, or unstable answering strategies. The existence of these anomalous responses is important to explain further because it can be used as a basis for evaluating the quality of student responses and the quality of the instruments used. According to (Boone et al., 2019), scalogram analysis in the Rasch model helps researchers identify response patterns that deviate from model expectations, thereby providing diagnostic information for improving instruments and learning processes.

4. CONCLUSION

This study shows that eight essay-type test items can be used to assess the critical thinking skills of junior high school students on the topic of temperature and heat. Rasch analysis indicates that most of the test items perform well and are capable of assessing students at various ability levels. The results also show that the test items do not function differently between male and female students. Wright's map and the scalogram provide additional information about differences in students' abilities and their response patterns. However, this study involved only 53 students from a single school, so the findings cannot yet be generalized to a broader population. In addition, the internal consistency of the instrument remains moderate. Future studies should involve students from multiple schools with a larger sample size and further improve the internal consistency of the instrument.

REFERENCES

Astini, N. K. S. (2022). Tantangan Implementasi Merdeka Belajar pada Era New Normal Covid-19 dan Era Society 5.0. *Journal LAMPUHYANG*, *13*(1), 1–17. https://e-journal.stkip-amlapura.ac.id

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2025). Development and validation of a complex true-false multiple-choice test for assessing critical thinking skills in physics. *Journal Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2553008

Bond, TG, & Fox, CM (2015). *Menerapkan model Rasch: Pengukuran fundamental dalam ilmu-ilmu humaniora* (edisi ke-3). Routledge.

Boone, W., Staver, J., & Yale, M. (2014). *Rasch Analysis in the Human Sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Carter, A. G., Müller, A., Gray, M., Bloxsome, D., Graham, K., Dooley, D., & Sweet, L. (2022). Critical thinking development in undergraduate midwifery students: an Australian validation study using Rasch analysis. *Journal BMC Pregnancy and Childbirth*, *22*(1), 1–12. https://doi.org/10.1186/s12884-022-05303-9

Dwikoranto., Lintangesukmanjaya, R. T., Setiani, & Rahyu. (2025). Analisis Model Rasch : Instrumen Penilaian Pemecahan Masalah Kreatif Dalam Pembelajaran Fisika. *Proceeding Seminar Nasional IPA XV*.

Febriyana S. (2023). Pengembangan Bahan Ajar Ipa Berbasis Gaya Belajar Dan Multiple Intelligences Untuk Meningkatkan Pemahaman Dan Keterampilan Berpikir Kritis Peserta Didik Pada Materi Suhu Kalor Dan Pemuaian. In *Tesis* (Issue 3, pp. 167–186). https://doi.org/http://repository.upi.edu/id/eprint/89984

Gleeson, J., Cutler, B., Rickinson, M., & Walsh, L. (2023). Rasch validation study. *Educational Assessment, Evaluation and Accountability*, *35*(2), 281–307. https://doi.org/10.1007/s11092-023-09404-7

Halpern, D., & Dunn, D. (2023). *Thought and Knowledge: An Introduction to Critical Thinking* (6th ed.). https://doi.org/doi.org/10.4324/9781003025412

Harahap, Y. A., & Zulhimma. (2026). Validitas dan Reliabilitas Instrumen Penilaian sebagai Dasar Analisis Hasil Tes serta Implementasi Umpan Balik, Program Perbaikan, dan Pengayaan dalam Pembelajaran. *RISET : Jurnal Ilmiah Multidisiplin Ilmu*, *1*(4), 1–13. https://doi.org/https://jurnalriset.com/index.php/riset/index

Ibrahim, N. L. (2025). Analisis Literatur Penggunaan Model Rasch untuk Validasi Instrumen Asesmen Pendidikan : Tren , Temuan , dan Implikasi. *Jurnal Inovasi Ilmu Pendidikan*, *3*(April), 1–15. https://doi.org/doi.org/10.55606/lencana.v3i2.5012

Linacre, J. M. (2010). Predicting Responses from Rasch Measures. *Journal of Applied Measurement*, *11*(1), 1–10.

Marfu'i, L. N. R., Ilfiandra, & Nurhudaya. (2019). The analysis of critical thinking skills test in social-problems for physics education students with Rasch Model. *Journal of Physics: Conference Series*, *1280*(5), 1–9. https://doi.org/10.1088/1742-6596/1280/5/052012

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research : Why measurement matters. *Journal Physical Review Physics Education Research*, *15*(2), 20111.

https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Sabekti, A. W., & Khoirunnisa, F. (2018). PENGGUNAAN RASCH MODEL UNTUK MENGEMBANGKAN INSTRUMEN PENGUKURAN KEMAMPUAN BERIKIR KRITIS SISWA PADA TOPIK IKATAN KIMIA. *Jurnal Zarah*, *6*(2), 68–75. https://doi.org/ojs.umrah.ac.id/index.php/zarah

Sujatmika, S. (2025). Applying the Rasch model to measure students ' critical thinking skills on the science topic of the human circulatory system. *Journal of Mathematics, Science and Technology Education*, *21*(4), 1–12. https://doi.org/doi.org/10.29333/ejmste/16221

Sumintono, B., & Widhiarso, W. (2014). Model, Rasch. In *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*.

Taber, K. S. (2018). The Use of Cronbach ' s Alpha When Developing and Reporting Research Instruments in Science Education Content courtesy of Springer Nature , terms of use apply . Rights reserved . *Journal Research in Science Education*, *48*, 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tennant, A., Mckenna, S. P., & Hagell, P. (2004). Application of Rasch Analysis in the Development and Application of Quality of Life Instruments. *Journal Value in Health*, *7*, S22–S26. https://doi.org/10.1111/j.1524-4733.2004.7s106.x

Tesio, L., Caronni, A., Kumbhare, D., Scarano, S., Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2024). Interpreting results from Rasch analysis 1 . The " most likely " measures coming from the model. *Journal Disability and Rehabilitation*, *46*(3), 591–603. https://doi.org/10.1080/09638288.2023.2169771

Tiruneh, D. T., de Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, *15*(4), 663–682. https://doi.org/10.1007/s10763-016-9723-0

Ubaidillah, M., Marwoto, P., Wiyanto, W., & Subali, B. (2023). How to Improve Critical Thinking in Physics Learning? A Systematic Literature Review. *Journal of Educational, Cultural and Psychological Studies*, *2023*(28), 161–187. https://doi.org/10.7358/ecps-2023-028-ubai

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299–321. https://doi.org/10.1080/00220272.2012.668938

Wulandari, D. T., Diani, R., & Noperi, H. (2025). Development of Diagnostic Instruments for Critical Thinking Skills of Students in Physics Learning on Socio-Scientific Issues (SSI). *Journal Corner of Education, Linguistics, and Literature*, *5*(1), 94–112. https://doi.org/10.54012/jcell.v5i1.534

Yew, E. H. J., Goh, K., Yew, E. H. J., & Goh, K. (2016). Problem-Based Learning : An Overview of its Process and Impact on Learning Problem-Based Learning : An Overview of its Process and Impact on Learning. *Health Professions Education*, *2*(2), 1–6. https://doi.org/10.1016/j.hpe.2016.01.004