



Exploring Students' Creative Thinking Skills: A Design of Reliable Instrument in term of Assessing Creative Thinking on Temperature and Heat Concepts

Ahmad Maqruf¹, Andi Suhandi^{1*}, Achmad Samsudin¹, Mustafa Sözbilir², Hasan Özgür Kapıcı³, Hadi Nasbey⁴, Eki Nugraha⁵

¹ Department of Physics Education, Universitas Pendidikan Indonesia, Indonesia

² Department of Mathematics and Science Education, Ataturk University, Türkiye

³ Department of Mathematics and Science Education, Bogazici University, Türkiye

⁴ Department of Physics Education, Universitas Negeri Jakarta, Indonesia

⁵ Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia

*Correspondence Author: andi_sb@upi.edu

ABSTRACT

This study aims to develop and validate an instrument for assessing creative thinking skills on the topic of temperature and heat among high school students. A quantitative research design was employed, with data analysis conducted using the Rasch model. The study involved 136 grade XI students from two high schools in West Bandung Regency, consisting of 51 male and 85 female participants. The Rasch analysis results indicated a Cronbach's alpha of 0.78 (good) and an item reliability value of 0.97 (excellent). However, minor revisions are required in the wording of three items to improve linguistic clarity. The percentage distribution of creative thinking indicators was as follows: fluency (70.34%), flexibility (68.57%), originality (60.13%), and elaboration (61.76%). These findings reveal that students' creative thinking is predominantly characterized by fluency, suggesting that they are relatively proficient at generating multiple ideas based on given problems, likely due to a good understanding of the material. Nevertheless, further instructional efforts are needed to foster higher levels of creative thinking, particularly in flexibility, originality, and elaboration.

Keywords: *creative thinking skills test, temperature and heat concepts, rasch analysis*

INTRODUCTION

Education in the 21st century has undergone significant transformation to meet the demands of rapid technological and societal change. These developments require continuous innovation to enhance the quality of education and ensure its relevance to current and future challenges (Permana et al., 2023; Ramadhanti & Azhar, 2022; Rani et al., 2019). The primary focus of today's educational process is the development of essential life skills both cognitive and interpersonal that are aligned with the competencies needed in this century (Haug & Mork, 2021; Suwistika et al., 2024). Within the cognitive domain, higher order thinking skills such as critical and creative thinking are of paramount importance (Álvarez Huerta et al., 2022). Consequently, learning designs should intentionally cultivate students' creative thinking abilities (Albar & Southcott, 2021). To effectively evaluate these skills, valid and reliable instruments are required to measure students' creative thinking development. Such instruments provide objective and accurate assessments that can help educators design more effective teaching strategies and monitor students' progress in developing

creative competencies an essential preparation for facing the increasingly complex challenges of the modern world.

Creative thinking has become a crucial skill for problem solving and innovation, influencing human progress and sustainability (Khoiri et al., 2019; Malik et al., 2019). It reflects the ability to organize thoughts to produce novel and useful ideas, typically evaluated through several dimensions such as fluency, flexibility, originality, and elaboration (Sahida & Zarvianti, 2019; Torrance, 2002). Fluency refers to the ability to generate a large number of relevant ideas quickly, flexibility involves proposing ideas from diverse perspectives (Wartono et al., 2018), originality represents the capacity to produce uncommon or unique ideas (Saregar et al., 2021), and elaboration denotes the ability to refine, expand, and implement ideas with detailed steps (Asriadi & Istiyono, 2020; Roshia & Hidayat, 2023). The development of an appropriate assessment instrument based on these indicators is therefore essential to accurately measure students' creative thinking skills.

One psychometric model that provides a robust framework for developing such instruments is the Rasch model. Initially introduced by Georg Rasch in 1960, this model examines item difficulty and individual ability simultaneously, allowing both to be placed on the same measurement scale (Astutik et al., 2020; Sa'diyah et al., 2021). The Rasch model applies probability theory to convert raw scores into interval data on a logit scale, providing a linear measure of student ability and item difficulty. It has become a powerful tool in educational measurement for ensuring construct validity and reliability in assessment development.

In physics education, creative thinking plays an important role in helping students understand and apply abstract scientific concepts. However, students often perceive physics particularly temperature and heat as a set of formulas to memorize rather than as concepts to understand (Kusairi, 2013; Nafi'ah et al., 2019). This misconception leads to difficulties in solving problems that require conceptual reasoning (Kamila et al., 2020; Sundari & Sarkity, 2021). Many students struggle with fundamental concepts such as thermal equilibrium and the distinction between heat and temperature (Kamar et al., 2016; Musa'adah & Kusairi, 2020). Because these phenomena are not directly observable, students' understanding is often shaped by their daily experiences rather than scientific reasoning (Aminudin et al., 2019; Budiarti et al., 2017).

The Indonesian government has introduced a curriculum aimed at developing students holistically across intellectual, creative, moral, and social dimensions to meet global challenges. According to the Ministry of Education and Culture, physics is not only a study of natural phenomena but also a means to foster self awareness as part of nature, emphasizing critical, creative, and ethical behavior in alignment with the Pancasila Student Profile (Kemendikbud, 2024). This approach aligns with UNESCO's four pillars of education learning to know, learning to do, learning to live together, and learning to be. Research by Trilling and Fadel (2012) also emphasizes that 21st century learners need skills such as communication, critical thinking, problem-solving, collaboration, and adaptability to thrive in diverse and technology-driven environments. Despite this, many students still lack competencies in technological literacy, project management, and leadership, underscoring the need to emphasize 21st century skills in science education.

Collaborative learning and authentic assessment are vital to achieving these educational goals. However, studies have shown that the development of valid instruments for measuring 21st century skills, especially creative thinking in physics, remains limited in Indonesia (Utari et al., 2020). Most assessment tools focus primarily on memory and low level cognitive processes (Nafi'ah et al., 2019; Herpiana et al., 2019). Teachers often face challenges in designing assessment instruments that evaluate creativity, leading to a lack of opportunities for students to engage in problem solving activities that stimulate innovative thinking. Consequently, there is a gap between the educational demand for fostering creativity and the availability of tools that can measure it effectively.

Therefore, this study aims to develop and validate a creative thinking skills test instrument focusing on temperature and heat concepts within the context of physics learning. The instrument is designed to assess students' ability to think fluently, flexibly, originally, and elaborately when solving physics related problems. The validation process, conducted using the Rasch model, is expected to ensure that the developed instrument meets the psychometric standards of validity and reliability required for assessing creative thinking in science education.

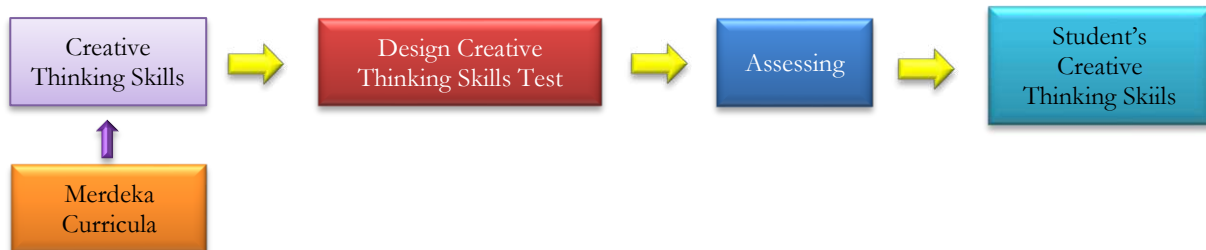


Figure 1. Framework for Creative Thinking Test on Heat and Temperature Concepts

METHODOLOGY

Design Study

This study employed a quantitative descriptive research design to systematically analyze and interpret data with the aim of producing accurate, reliable, and high-quality findings (Duckett, 2021). The research was conducted through four stages: developing the creative thinking test instrument, administering it to students, analyzing the data, and interpreting the results (Amiruddin et al., 2023). The instrument was constructed based on four indicators fluency, flexibility, originality, and elaboration and was distributed via Google Forms to high school students in West Bandung Regency. Data were processed using Excel for preliminary analysis, WINSTEPS for Rasch model analysis, and Notepad for data coding. The Rasch model was employed to determine item difficulty, person ability, and instrument reliability. The results were then interpreted based on classroom conditions to ensure contextual and educational relevance.

Participant

The participants in this study consisted of 136 students (51 males and 85 females) from two senior high schools located in West Bandung Regency, West Java, Indonesia. All participants were Grade XI students enrolled under the Merdeka Curriculum, which emphasizes student centered learning and the development of 21st century competencies. The selection of participants was based on their enrollment in physics courses covering the topic of temperature and heat. The research sites were chosen to represent schools with similar academic characteristics and learning environments. The geographical location of the study area is presented in Figure 2.

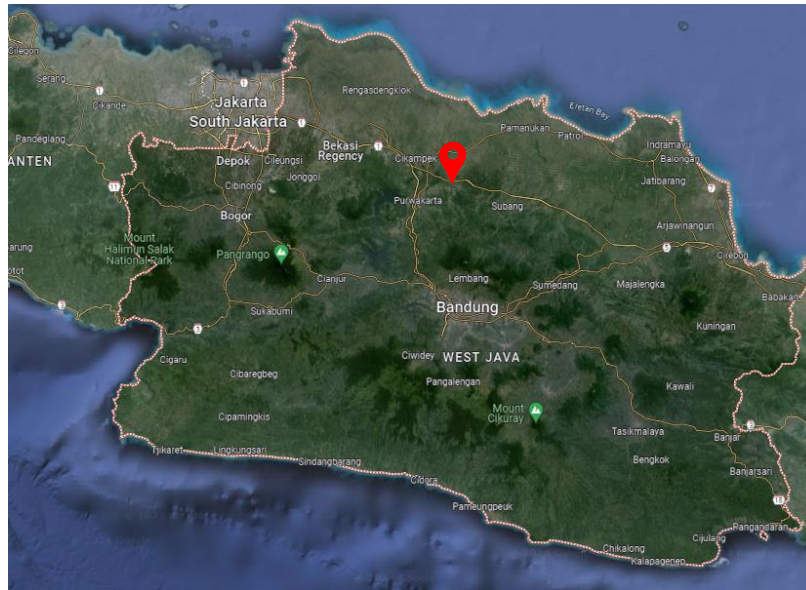


Figure 2. Distribution of Participant

Instrument

This study utilized six descriptive test items designed to assess students' creative thinking skills in the context of physics learning. The questions focused on three key concepts: heat and temperature change, heat and physical change (expansion), and heat transfer. Each item was developed to measure one or more of the four dimensions of creative thinking fluency, flexibility, originality, and elaboration. The instrument underwent expert validation by a panel of five experts, consisting of two university physics lecturers and three high school physics teachers. These experts were selected based on their strong understanding of physics concepts, extensive teaching experience, and proven competence in evaluating educational materials in accordance with national educational standards. The validation process assessed several criteria, including the alignment of test items with the indicators of creative thinking, the clarity and readability of the language used, and the relevance of the items to the concepts of temperature and heat. Subsequently, the validity and reliability of the test items were examined using Rasch model analysis to ensure psychometric soundness and measurement precision. The finalized test items used in this study are presented in Figure 3.

Mr. Rahmat always brews coffee every morning before going to work. However, today he woke up late, and he only had 10 minutes left before he had to leave for work. As usual, he put a kettle of water (2liters) on the stove to boil water for making coffee. However, after 10 minutes, the water had not boiled. Finally, Mr. Rahmat did not brew coffee today. Help Mr. Rahmat find a solution to his problem for the future.

- A. List as many factors as possible that could cause the water not to boil within 10 minutes!
- B. List as many creative ideas as possible to solve Mr. Rahmat's problem!
- C. Based on your analysis of the mistakes made by Mr. Rahmat, design a coffee maker that can boil water in less than 10 minutes!
- D. Describe the advantages of the design you have made!

Figure 3. Examples of Creative Thinking Skills Test

Data Analysis

The data analyzed in this study consisted of expert validation results and students' responses to the test items. The analysis process was carried out using the Rasch model, which allows for the examination of both item and person parameters on a common measurement scale. The Rasch analysis provided detailed information regarding item validity, respondent reliability, item reliability, as well as logit values for both items and respondents. The validity of the Creative Thinking Skills Instrument on the topic of temperature and heat serves as a critical indicator of whether the developed cognitive assessment tool effectively measures students' creative thinking abilities in this domain. The interpretation of the analysis results followed the guidelines proposed by Sumintono and Widhiarso (2014), focusing on key psychometric indicators such as item fit, person fit, and reliability indices. The summary of instrument validity findings is presented in Table 1.

Table 1. Interpretation of Instrument Test Validity

Interpretation	Raw variance explained by measure
Unfulfilled	$V < 20\%$
Fulfilled	$20\% \leq V \leq 40\%$
In Accordance	$40\% < V \leq 60\%$
Special	$V > 60\%$

Reliability items are used to determine the level of confidence that an instrument will produce the same results when used repeatedly, or to determine the consistency of test instruments used in research. The interpretation of the results is presented as follows (Sumintono & Widhiarso, 2014) person and item reliability (Table 2).

Tabel 2. Interpretation of Item and Person Reliability

Interpretation	Value of Person Reliability and Item Reliability
Weak	$r < 0.67$
Moderate	$0.67 \leq r < 0.80$
Good	$0.80 \leq r < 0.90$
Very Good	$0.90 \leq r < 0.94$
Excellent	$r \geq 0.94$

Furthermore, Logit items are used to indicate the difficulty and ease of a test question instrument. In addition, person logit is used to see the ability of students' creative thinking skills in response to test instruments. The interpretation of the results is presented as follows (Sumintono & Widhiarso, 2014) item and person logit (Table3).

Table 3. Interpretation of Item and Person Logit

Interpretation	Value of Item Logit (A)	Interpretation	Value of Person Logit (B)
Very Easy	$A < -0.82$	Low	$B < 0.96$
Easy	$-0.82 \leq A \leq 0.00$	Moderate	$0.96 \leq B < 1.78$
Difficult	$0.00 < A \leq 0.82$	High	$B > 1.78$
Very Difficult	$A > 0.82$		

RESULT AND DISCUSSION

The results of this study are presented in several subsections to provide a detailed explanation and to address the research questions comprehensively. Data analysis and interpretation were conducted using the Rasch model, which minimizes human error and provides objective and consistent measurement results. The Rasch analysis allows for a more precise evaluation of test items and respondent performance, ensuring that the developed instrument meets psychometric standards for educational assessment. One of the key evaluations in determining the validity of the developed test instrument is unidimensionality analysis (Darmana et al., 2021; Nurdini et al., 2020). This analysis is used to confirm whether the instrument measures a single underlying construct in this case, students' creative thinking skills. A unidimensional instrument ensures that all items contribute meaningfully to assessing the same latent trait rather than multiple unrelated abilities. The results of the unidimensionality analysis for the Creative Thinking Skills Instrument on temperature and heat concepts are presented in Figure 4.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units			
	Eigenvalue	Observed	Expected
Total raw variance in observations =	23.6524	100.0%	100.0%
Raw variance explained by measures =	9.6524	40.8%	40.9%
Raw variance explained by persons =	4.8843	20.7%	20.7%
Raw Variance explained by items =	4.7681	20.2%	20.2%
Raw unexplained variance (total) =	14.0000	59.2%	100.0%
Unexplned variance in 1st contrast =	2.8364	12.0%	20.3%
Unexplned variance in 2nd contrast =	1.8207	7.7%	13.0%
Unexplned variance in 3rd contrast =	1.7732	7.5%	12.7%
Unexplned variance in 4th contrast =	1.5401	6.5%	11.0%
Unexplned variance in 5th contrast =	1.3020	5.5%	9.3%

Figure 4. Instrument Validity Results

Based on Figure 4, the results of the validity measurement indicate that the raw variance explained by the measures is 40.8% (orange box), which falls within the acceptable range of 40% to 60% (see Table 1). This suggests that the instrument meets the criterion for good unidimensionality and can therefore be considered appropriate for further research. In the Rasch model, the assessment of unidimensionality is conducted using Principal Component Analysis (PCA) of residuals, which examines the extent to which the developed instrument measures a single latent construct as intended (Amiruddin et al., 2023; Hagell, 2014; Soeharto, 2021).

Another important indicator in this analysis is the unexplained variance, particularly the first contrast, which represents the strength of the secondary dimension within the data. The acceptable threshold for this value is below 15%, indicating minimal multidimensionality. The obtained value of 12.0% (red box) meets this criterion, and the values of subsequent contrasts are also below 15%. These findings confirm that the instrument demonstrates satisfactory construct validity, meaning that all items consistently measure students' creative thinking skills within the domain of temperature and heat concepts. Furthermore, the item and person analysis provides complementary information about the reliability of the questionnaire, including item reliability, person reliability, and the relationship between item person measures on the developed instrument. The detailed results of these analyses are presented in Figure 5.

SUMMARY OF 136 MEASURED Person								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	26.7	14.0	.96	.39	1.01	-.18	.99	-.20
SEM	.5	.0	.07	.00	.06	.13	.06	.13
P.SD	5.7	.0	.82	.02	.65	1.52	.66	1.50
S.SD	5.7	.0	.83	.02	.65	1.53	.66	1.50
MAX.	37.0	14.0	2.65	.49	3.76	4.71	3.67	4.49
MIN.	14.0	14.0	-.90	.37	.19	-3.61	.19	-3.48
REAL RMSE	.43	TRUE SD	.70	SEPARATION	1.64	Person RELIABILITY	.73	
MODEL RMSE	.39	TRUE SD	.73	SEPARATION	1.87	Person RELIABILITY	.78	
S.E. OF Person MEAN = .07								
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00								
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .78 SEM = 2.70								
STANDARDIZED (50 ITEM) RELIABILITY = .93								
SUMMARY OF 14 MEASURED Item								
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	259.0	136.0	.00	.12	1.00	-.09	.99	-.16
SEM	15.2	.0	.23	.00	.06	.61	.06	.56
P.SD	54.9	.0	.82	.01	.22	2.40	.22	2.02
S.SD	57.0	.0	.85	.01	.24	2.27	.22	2.09
MAX.	356.0	136.0	1.57	.16	1.43	3.52	1.31	2.42
MIN.	153.0	136.0	-1.61	.12	.46	-5.96	.48	-5.48
REAL RMSE	.13	TRUE SD	.81	SEPARATION	6.22	Item RELIABILITY	.97	
MODEL RMSE	.12	TRUE SD	.81	SEPARATION	6.52	Item RELIABILITY	.98	
S.E. OF Item MEAN = .23								
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00								
Global statistics: please see Table 44.								
UMEAN=.0000 USCALE=1.0000								

Figure 5. Item person Reliability of The Instrument

Figure 5 presents detailed information regarding the relationship between items and respondents (persons) based on Cronbach's alpha values. The person reliability value (red box) is 0.73, which falls under the fair category, while the item reliability value (yellow box) is 0.97, indicating an excellent level of reliability. These results imply that the students' response consistency is acceptable, and the developed test items demonstrate very high quality. Thus, based on the Cronbach's alpha coefficient, the overall relationship between item and person reliability is well established (Robinson, 2018; Surucu & Maslakci, 2020; Taber, 2018).

The overall reliability index of 0.78 (red box) is categorized as good, indicating that the instrument is sufficiently consistent for assessing students' creative thinking skills. This also suggests that students' creative thinking abilities exhibit notable variability, allowing for differentiation in their performance levels. Consequently, the instrument is suitable for profiling students' creative thinking skills within the context of temperature and heat concepts. Furthermore, this analysis reveals both the distribution of students' creative thinking ability levels and the difficulty level of each test item, reinforcing the validity and reliability results obtained previously (Cheung et al., 2023; Tennant & Conaghan, 2007; White & Ronfeldt, 2024). Therefore, a more detailed examination of item and person logit values is required to clarify the findings and to identify how well the test items represent the intended measurement construct. The results of this analysis are presented in Figure 6.

Item STATISTICS: MEASURE ORDER												ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PTMEASUR-AL CORR.	EXACT MATCH EXP% OBS% EXP%	Item	Person							
12	153	136	1.57	.131	.71	-2.63	.72	-2.46	.44	.47	66.9	55.9	Q4D	69	37	14	2.65	.491	.47	-1.30	.58	-.64	.40	.60	.40	78.6	67.9	069F	
4	183	136	1.10	.121	.46	-5.96	.48	-5.48	.53	.49	73.5	53.0	Q1D	5	36	14	2.42	.46	1.16	.49	1.27	.65	.24	.43	.57	1.1	64.0	005F	
11	210	136	.71	.12	.89	-.98	.90	-.90	.57	.51	54.4	49.5	Q4C	22	36	14	2.42	.46	.82	-.32	.62	-.64	.71	.43	.64	3.3	64.0	022F	
1	214	136	.65	.12	1.07	-.64	1.07	.61	.41	.51	47.1	49.3	Q1A	35	36	14	2.42	.46	1.24	.68	1.83	1.42	.13	.43	.57	1.1	64.0	035F	
3	225	136	.50	.12	.93	-.64	1.00	.03	.48	.52	54.4	47.9	Q1C	104	36	14	2.42	.46	.78	-.43	.67	-.52	.58	.43	.78	6.4	104F		
10	249	136	.17	.12	.99	-.03	1.01	.12	.58	.53	48.5	46.7	Q4B	9	35	14	2.22	.44	.51	-1.42	.60	-.80	.52	.45	.71	4.2	009F		
13	260	136	.01	.12	1.15	1.37	1.22	1.90	.50	.53	39.7	45.8	Q5	17	35	14	2.22	.44	.55	-1.26	.57	-.88	.51	.45	.57	1.1	62.0	017M	
2	268	136	-.10	.12	1.09	.85	1.08	.75	.60	.53	30.9	46.0	Q1B	123	35	14	2.22	.44	.63	-.96	.52	-1.04	.80	.45	.85	7.2	021F		
8	280	136	-.27	.12	1.12	1.14	1.15	1.33	.33	.52	53.7	47.4	Q3B	79	17	14	-.41	.39	1.25	.73	1.35	.95	-.26	.48	.64	3.3	55.2	079F	
9	298	136	-.54	.12	.91	-.78	.87	-1.08	.61	.52	47.1	50.4	Q4A	71	16	14	-.57	.40	2.91	3.46	2.90	3.43	.68	.48	.7	1.1	56.7	071F	
14	301	136	-.58	.12	1.43	3.52	1.31	2.42	.59	.51	48.5	50.6	Q6	72	16	14	-.57	.40	2.91	3.46	2.90	3.43	.68	.48	.7	1.1	56.7	072F	
7	307	136	-.68	.13	1.21	1.82	1.29	2.16	.35	.51	49.3	51.5	Q3A	75	16	14	-.57	.40	2.91	3.46	2.90	3.43	.68	.48	.7	1.1	56.7	075F	
6	322	136	-.93	.13	.89	-.90	.80	-1.51	.65	.49	63.2	54.8	Q2B	82	16	14	-.57	.40	.54	-1.30	.51	-1.42	.09	.48	.71	4.2	082F		
5	356	136	-1.61	.16	1.20	1.37	.97	-.13	.46	.43	66.9	68.4	Q2A	87	16	14	-.57	.40	1.10	.37	1.08	.33	.09	.48	.57	1.1	56.7	087F	
MEAN	259.0	136.0	.00	.12	1.00	-.09	.99	-.16			53.2	51.2		93	16	14	-.57	.40	.54	-1.30	.51	-1.42	.09	.48	.71	4.2	093M		
P.S.D	54.9	.0	.82	.01	.23	2.19	.22	2.02			11.0	5.6		74	14	14	-.90	.42	3.55	4.11	3.42	4.02	.56	.46	.7	1.1	59.0	074F	
MEAN	25.7	14.0	.96								39.1	18.1		MEAN	26.7	14.0	.96	.39	1.01	-.18	.99	-.20						53.2	51.2
P.S.D	5.7	.0	.82								11.0	5.6		P.S.D	5.7	.0	.82	.02	.65	1.52	.66	1.50						17.6	4.8

Figure 6. Item Person Logit Values

Based on the logit values obtained according to Figure 5, the students' creative thinking skills and the difficulty level of the questions can be analyzed in depth (see Table 3). For item logits, the most difficult question is Q4D with a logit of 1.57 (very difficult), while the easiest question is Q2A with a logit of -1.61 (easy). Although item Q4D is very difficult, a quarter of respondents with high creative thinking skills were able to answer it well. Then, the student with the highest creative thinking ability is 069F with a logit value of 2.65 (high), while the student with the lowest creative thinking ability is 074F with a logit value of -0.90 (low). This indicates that the reliability of the questions also affects the students' creative thinking skills. The mapping of students' creative thinking skills to the questions is shown in Figure 7.

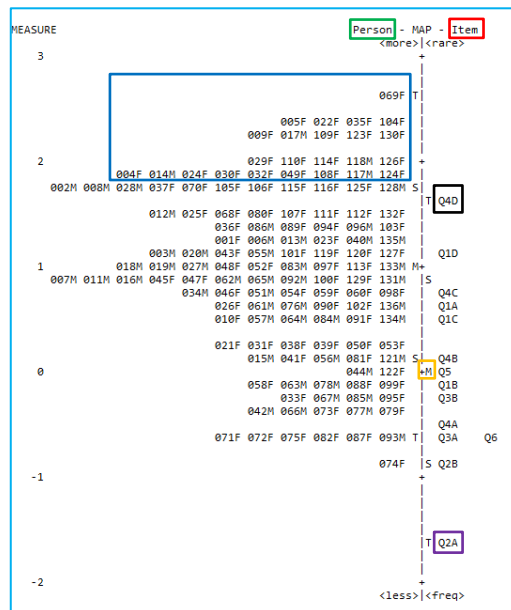


Figure 7. The Output of Variable (Wright) Maps

Based on Figure 7, the distribution of students' creative thinking skills and the difficulty levels of the test items can be clearly observed. Students represented within the blue box demonstrate very high proficiency, indicating their ability to correctly answer all items. In contrast, item Q4D (black box) is identified as the most difficult question within the instrument. Meanwhile, the student labeled 074F falls into the very low ability category, as this participant only responded correctly to items Q2A and Q2B, which were classified as the easiest items. Overall, students' creative thinking skills were assessed through four aspects: fluency, flexibility, originality, and elaboration.

Figure 7 also visualizes the interaction between items (red boxes) and persons (green boxes). The Item section lists each question (coded as Q1–Q6), while the Person section includes participant identifiers, distinguished by gender through the codes M (male) and F (female). The response distribution follows an approximately normal curve, with most students clustering around the median level of ability (yellow box). Students positioned in the blue box represent those with the highest creative thinking ability, while those in the red box represent individuals with limited ability to respond accurately to the given questions.

The Wright map (variable map) output provides an effective visualization of the probability of each student successfully answering a given item, thus allowing researchers to interpret both ability and item difficulty within the same measurement continuum (Darman et al., 2024; Prasetya & Pratama, 2023; Sumintono, 2018). Furthermore, to examine potential gender related bias in the developed instrument, a Differential Item Functioning (DIF) analysis was conducted, as illustrated in Figure 8.

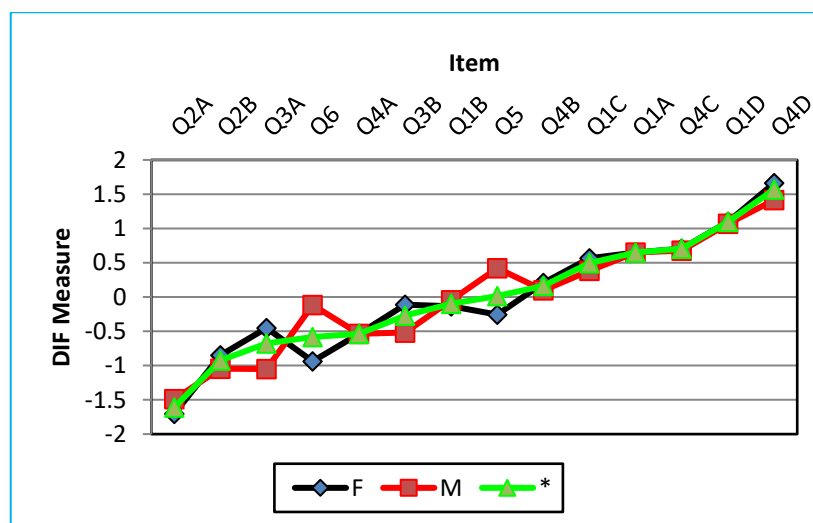


Figure 8. DIF Between Gender Differences in Answering Questions

Differential Item Functioning (DIF) refers to differences in the probability of correctly answering an item between two distinct groups (Kucam & Gülleroğlu, 2023; Samsudin et al., 2023). In Figure 8, the DIF measure (Diff.) represents the level of item difficulty where a higher value indicates a more challenging item. The black line represents the female group (F), while the red line represents the male group (M). The sequence of questions is displayed along the x-axis. The increasing trend in the graph illustrates that the overall difficulty of the items shows a positive progression, indicating that the instrument effectively differentiates between varying levels of student ability.

However, gender bias was identified in several items where the F or M curves deviate above or below the green reference line (normal line). For instance, in item Q3A, the blue line appears above both the green and red lines, suggesting that female students are more likely to answer this item correctly compared to males. Conversely, for items Q5 and Q6, the red line appears above the green and blue lines, indicating that these questions are easier for male students. Therefore, minor revisions are recommended to reduce gender bias and enhance fairness across items.

This analysis also contributes to identifying outlier items or individuals that do not align with the expected measurement model. Three main indicators are used to assess item and person fit: the Mean Square (MNSQ), which is considered acceptable when $0.5 < MNSQ < 1.5$; the Z-Standardized (ZSTD) value, which is acceptable when $-2.0 < ZSTD < +2.0$; and the Point Measure

Correlation (Pt Mean Corr), which is acceptable when $0.4 < Pt\ Mean\ Corr < 0.85$ (Sumintono & Widhiarso, 2014). The results of this fit analysis are presented in Figure 9.

Item STATISTICS: MISFIT ORDER													
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	TMEAS CORR.	R-AL EXP.	EXACT OBS%	MATCH EXP%	Item
14	301	136	-.58	.12	1.43	3.52	.31	2.42	.59	.51	48.5	50.6	Q6
7	307	136	-.68	.13	1.21	1.82	.29	2.16	.35	.51	49.3	51.5	Q3A
13	260	136	.01	.12	1.15	1.37	.22	1.90	.50	.53	39.7	45.8	Q5
5	356	136	-1.61	.16	1.20	1.37	.97	-1.13	.46	.43	66.9	68.4	Q2A
8	280	136	-.27	.12	1.12	1.14	.15	1.33	.33	.52	53.7	47.4	Q3B
2	268	136	-.10	.12	1.09	.85	.08	.75	.60	.53	30.9	46.0	Q1B
1	214	136	.65	.12	1.07	.64	.07	.61	.41	.51	47.1	49.3	Q1A
10	249	136	.17	.12	.99	-.03	.01	.12	.58	.53	48.5	46.7	Q4B
3	225	136	.50	.12	.93	-.64	.00	.03	.48	.52	54.4	47.9	Q1C
9	298	136	-.54	.12	.91	-.78	.87	1.08	.61	.52	47.1	50.4	Q4A
11	210	136	-.71	.12	.89	-.98	.90	-1.90	.57	.51	54.4	49.5	Q4C
6	322	136	-.93	.13	.89	-.90	.80	1.51	.65	.49	63.2	54.8	Q2B
12	153	136	1.57	.13	.71	-2.63	.72	2.46	.44	.47	66.9	55.9	Q4D
4	183	136	1.10	.12	.46	-5.96	.48	5.48	.53	.49	73.5	53.0	Q1D
MEAN	259.0	136.0	.00	.12	1.00	-.09	.99	-1.16			53.2	51.2	
P.SD	54.9	.0	.82	.01	.23	2.19	.22	2.02			11.0	5.6	

Figure 9. Item Fit Order

Figure 9 presents the item fit analysis for the Creative Thinking Skills Instrument in measuring students' creative thinking abilities. Based on the ZSTD criterion, several items are classified as misfitting, including Q6, Q3A, Q4D, and Q1D. Specifically, items Q3A and Q6 exceed the upper threshold of 2.0, while items Q4D and Q1D fall below the lower threshold, with values of -2.46 and -5.48 , respectively. Nevertheless, when evaluated using MNSQ and Point Measure Correlation (Pt Mean Corr) criteria, all items meet the acceptable fit standards, indicating that the overall instrument remains psychometrically acceptable and suitable for assessing students' creative thinking skills.

To further interpret students' performance, the percentage of achievement in each dimension of creative thinking was calculated, providing a more comprehensive view of students' ability distribution. The results for the percentage of each category of creative thinking skills fluency, flexibility, originality, and elaboration related to temperature and heat concepts are presented in Figure 10. This analysis allows for a clearer understanding of the strengths and weaknesses of students' creative thinking profiles and supports the validity and reliability findings discussed earlier.

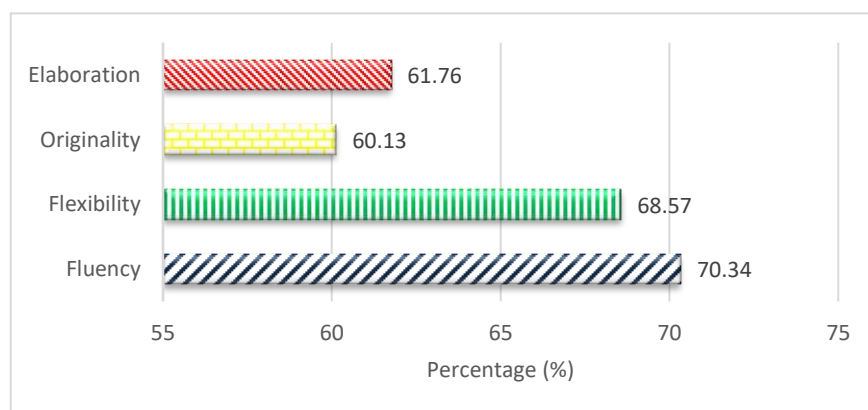


Figure 10. Percentage of Each West Bandung Regency Student Creative Thinking Skills Category

Figure 10 illustrates the percentage distribution of students' creative thinking skills across four categories in West Bandung Regency. The percentages from highest to lowest are: fluency (70.34%), flexibility (68.57%), elaboration (61.76%), and originality (60.13%). These results indicate that fluency is the strongest aspect of students' creative thinking regarding temperature and heat concepts. This suggests that physics students in West Bandung Regency are proficient at generating multiple ideas in response to the given questions, likely due to a solid understanding of the

underlying material. These findings are consistent with previous studies, which reported that students' creative thinking skills tend to be dominated by fluency (Nurdiana, 2020; Sa'diyah et al., 2021; Syukri et al., 2022).

However, some student responses still rely heavily on theoretical knowledge provided by teachers and show limited connection to experiential observations of their environment. To address this, learning activities on temperature and heat material could be enhanced by incorporating a 21st century skills approach, particularly emphasizing creative thinking development. The results of this study also confirm that the developed instrument is valid and reliable, effectively capturing the profile of students' creative thinking skills. Nevertheless, minor revisions are necessary for certain items that demonstrated suboptimal performance to further improve accuracy and effectiveness. Conducting additional evaluations and retesting will ensure broader applicability and provide deeper insights into students' creative thinking abilities.

It should be noted that this study relies solely on quantitative data derived from percentages of creative thinking skills, without incorporating qualitative data that could offer richer insights into how students develop and apply their creativity. Additionally, the sample was limited to students from West Bandung Regency, which may restrict the generalizability of the findings to other regions or educational levels. Therefore, future research is recommended to include more diverse samples and to integrate qualitative data collection methods, enabling a more comprehensive understanding of students' creative thinking skills across different contexts.

CONCLUSION

This study successfully developed and validated a creative thinking skills test instrument on the topic of temperature and heat for high school students. The Rasch analysis results indicate that the instrument is valid for measuring a single dimension of creative thinking, encompassing fluency, flexibility, originality, and elaboration within the physics context. The analysis also revealed a clear relationship between items and respondents, with a Cronbach's alpha of 0.78, classified as good, indicating satisfactory internal consistency. The instrument demonstrated excellent reliability, as evidenced by consistent item fit across the test. Among the participants, the highest creative thinking ability was observed in student 069F, with the highest logit value exceeding the most difficult item, Q4D, while the lowest ability was recorded for student 072F, corresponding to the easiest item, Q2A. Some gender bias was identified in items Q3A, Q5, and Q6, suggesting that minor revisions are required to enhance fairness. It is recommended that this instrument be further developed and refined through testing on a larger and more diverse sample to ensure broader generalizability. Physics teachers can utilize this tool to identify, monitor, and map students' creative thinking skills, enabling more targeted instructional strategies. Further research is also needed to examine the validity and reliability of this instrument across other physics concepts and at different educational levels to expand its applicability.

ACKNOWLEDGMENTS

The authors would like to sincerely express their gratitude to the Riau Provincial Government, especially the Riau Provincial Education Office, for providing financial support and opportunities through the Master's Degree Scholarship Program at Universitas Pendidikan Indonesia. Their support was instrumental in facilitating the completion of this research.

REFERENCES

Albar, S. B., & Southcott, J. E. (2021). Problem and Project Based Learning Through an Investigation Lesson: Significant Gains in Creative Thinking Behaviour Within The Australian

- Foundation (Preparatory) Classroom. *Thinking Skills and Creativity*, 41(May 2019), 100853. <https://doi.org/10.1016/j.tsc.2021.100853>
- Álvarez Huerta, P., Muela, A., & Larrea, I. (2022). Disposition Toward Critical Thinking and Creative Confidence Beliefs in Higher Education Students: The Mediating Role of Openness to Diversity and Challenge. *Thinking Skills and Creativity*, 43(December 2021). <https://doi.org/10.1016/j.tsc.2022.101003>
- Aminudin, A. H., Rusdiana, D., Samsudin, A., Hasanah, L., & Maknun, J. (2019). Measuring Critical Thinking Skills of 11th Grade Students on Temperature and Heat. *Journal of Physics: Conference Series*, 1280(5). <https://doi.org/10.1088/1742-6596/1280/5/052062>
- Amiruddin, M. Z. Bin, Samsudin, A., Suhandi, A., Kaniawati, I., COŞTU, B., & Aminuddin, Adam Hadiana Kuniawan, F. (2023). Validity and Reliability of the Global Warming Instrument: A Pilot Study Using Rasch Model Analysis. *Jurnal Pendidikan MIPA*, Vol 24, No 4 (2023): *Jurnal Pendidikan MIPA*, 912–922. <https://jurnal.fkip.unila.ac.id/index.php/jpmipa/article/view/29266/pdf>
- Asriadi, M., & Istiyono, E. (2020). Exploration of Creative Thinking Skills of Students in Physics Learning. *Journal of Educational Science and Technology (EST)*, 6(2), 151–158. <https://doi.org/10.26858/est.v6i2.12737>
- Astutik, S., Mahardika, I. K., Indrawati, Sudarti, & Supeno. (2020). HOTS Student Worksheet to Identification of Scientific Creativity Skill, Critical Thinking Skill and Creative Thinking Skill in Physics Learning. *Journal of Physics: Conference Series*, 1465(1). <https://doi.org/10.1088/1742-6596/1465/1/012075>
- Atmojo, I. R. W., & Sajidan. (2020). Effectiveness of CEL-badis Learning Model on Students' Creative Thinking Skills: Case on The Topic of Simple Food Biotechnology. *International Journal of Instruction*, 13(3), 329–342. <https://doi.org/10.29333/iji.2020.13323a>
- Budiarti, I. S., Suparmi, A., Sarwanto, & Harjana. (2017). Analyzes of Students' Higher-Order Thinking Skills of Heat and Temperature Concept. *Journal of Physics: Conference Series*, 909(1). <https://doi.org/10.1088/1742-6596/909/1/012055>
- Cheung, G. W., Cooper Thomas, H. D., Lau, R. S., & Wang, L. C. (2023). Reporting Reliability, Convergent and Discriminant Validity with Structural Equation Modeling: A Review and Best Practice Recommendations. In *Asia Pacific Journal of Management* (Vol. 41, Issue 2). Springer US. <https://doi.org/10.1007/s10490-023-09871-y>
- Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and Validation of Scientific Inquiry Literacy Instrument (SILI) Using Rasch Measurement Model. *Education Sciences*, 14(3). <https://doi.org/10.3390/educsci14030322>
- Darmana, A., Sutiani, A., Nasution, H. A., Ismanisa*, I., & Nurhaswinda, N. (2021). Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments. *Jurnal Pendidikan Sains Indonesia*, 9(3), 329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Duckett, L. J. (2021). Quantitative Research Excellence: Study Design and Reliable and Valid Measurement of Variables. *Journal of Human Lactation*, 37(3), 456–463. <https://doi.org/10.1177/08903344211019285>
- Hagell, P. (2014). Testing Rating Scale Unidimensionality Using the Principal Component Analysis (PCA) Test Protocol with the Rasch Model: The Primacy of Theory over Statistics. *Open Journal of Statistics*, 04(06), 456–465. <https://doi.org/10.4236/ojs.2014.46044>
- Haug, B. S., & Mork, S. M. (2021). Taking 21st Century Skills From Vision to Classroom: What Teachers Highlight as Supportive Professional Development in The Light of New Demands From Educational Reforms. *Teaching and Teacher Education*, 100, 103286.

<https://doi.org/10.1016/j.tate.2021.103286>

- Heliawati, L., Afakillah, I. I., & Pursitasari, I. D. (2021). Creative Problem Solving Learning Through Open Ended Experiment for Students' Understanding and Scientific Work Using Online Learning. *International Journal of Instruction*, 14(4), 321–336. <https://doi.org/10.29333/iji.2021.14419a>
- Herpiana, R., Rosidin, U., & Abdurrahman, A. (2019). Development of Instruments to Train Critical and Creative Thinking Skills in Physics Assessment for High School Students' Learning. *Journal of Physics: Conference Series*, 1155(1). <https://doi.org/10.1088/1742-6596/1155/1/012046>
- Istiyono, E. dkk. (2018). Pengembangan Tes. *Cakrawala Pendidikan*, 190–200.
- Kamar, N., Kusairi, S., & Zubaidah, S. (2016). Pengembangan Asesmen Formatif dan Remediasi Berbasis Komputer pada Pokok Bahasan Subu dan Kalor SMP Kelas VII 1, 2, 3. 4(2), 66–76.
- Kamila, N., Budi Prastowo, S. H., & Nuraini, L. (2020). Penerapan Model Application of Step Instruction and Elaboration (APPOSITE) untuk Meningkatkan Hasil Belajar Kognitif Siswa SMA pada Materi Suhu dan Kalor. *Journal of Natural Science and Integration*, 3(2), 163. <https://doi.org/10.24014/jnsi.v3i2.9348>
- Kemendikbud. (2024). *Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada Kurikulum Merdeka* (Issue 021).
- Khoiri, A., Nulngafan, N., Sunarno, W., & Sajidan, S. (2019). How is Students' Creative Thinking Skills? An Ethnoscience Learning Implementation. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 8(2), 153–163. <https://doi.org/10.24042/jipfalbiruni.v0i0.4559>
- Kucam, E., & Gülleroğlu, H. D. (2023). Examination of Differential Item Functioning in PISA 2018 Mathematics Literacy Test with Different Methods. *Journal of Measurement and Evaluation in Education and Psychology*, 14(2), 128–135. <https://doi.org/10.21031/epod.1122857>
- Kusairi, S. (2013). Analisis Asesmen Formatif Fisika SMA berbantuan Komputer. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 16(3), 68–87. <https://doi.org/10.21831/pep.v16i0.1106>
- Malik, A., Nuraeni, Y., Samsudin, A., & Sutarno, S. (2019). Creative Thinking Skills of Students on Harmonic Vibration using Model Student Facilitator and Explaining (SFAE). *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, 8(1), 77–88. <https://doi.org/10.24042/jipfalbiruni.v8i1.3056>
- Musa'adah, R., & Kusairi, S. (2020). Analisis Penguasaan Konsep Siswa Materi Suhu Kalor pada Pembelajaran Cognitive Apprenticeship (CA) disertai Formative E-Assessment. *JRPF (Jurnal Riset Pendidikan Fisika)*, 4(2), 85–90.
- Nafi'ah, Z., Mustikasari, V. R., & Pratiwi, N. (2019). Pengembangan Instrumen Tes Two Tier Multiple Choice untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Peserta Didik SMP pada Materi Suhu dan Kalor. *JIPVA (Jurnal Pendidikan IPA Veteran)*, 3(2), 115. <https://doi.org/10.31331/jipva.v3i2.777>
- Nurdiana, H. (2020). Creative Thinking Skills Profile of Junior High School Students in Science Learning. *Journal of Physics: Conference Series*, 1567(2). <https://doi.org/10.1088/1742-6596/1567/2/022049>
- Nurdini, N., Suhandi, A., Ramalis, T., Samsudin, A., Fratiwi, N. J., & Costu, B. (2020). Developing Multitier Instrument of Fluids Concepts (MIFO) to Measure Student's Conception: A Rasch Analysis Approach. *Journal of Advanced Research in Dynamical and Control Systems*, 12(6), 3069–3083. <https://doi.org/10.5373/JARDCS/V12I6/S20201273>
- Permana, N. D., Lestari, I., Susanty Harahap, F. D., Azhar, A., & Defianti, A. (2023). Project Based Learning (PjBL) Model with STEAM Approach: Its Impact on Students Creative Thinking

- Skills on Energy in Living System Topic. *Journal of Natural Science and Integration*, 6(2), 186. <https://doi.org/10.24014/jnsi.v6i2.25791>
- Prasetya, W. A., & Pratama, A. T. (2023). Item Quality Analysis Using The Rasch Model to Measure Critical Thinking Ability in The Material of The Human Digestive System of Biology Subject in High School. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 27(1), 76–91. <https://doi.org/10.21831/pep.v27i1.58873>
- Ramadhanti, R. N., & Azhar. (2022). E-Module Validation Based on Contextual Teaching and Learning for High School Optical Instruments. *Journal of Natural Science and Integration P-ISSN*, 5(2), 237–246. <https://doi.org/10.24014/jnsi.v5i2.16982>
- Rani, S. A., Mundilarto, Warsono, & Dwandaru, W. S. B. (2019). Physics Virtual Laboratory: An Innovative Media in 21st Century Learning. *Journal of Physics: Conference Series*, 1321(2). <https://doi.org/10.1088/1742-6596/1321/2/022026>
- Robinson, M. A. (2018). Using Multi-Item Psychometric Scales for Research and Practice in Human Resource Management. *Human Resource Management*, 57(3), 739–750. <https://doi.org/10.1002/hrm.21852>
- Rosha, J. M., & Hidayat, A. (2023). Analysis of Creative Thinking Skill Instrument Test (CreaTSIT) on Renewable Energy Topic for Senior High School Student using Rasch Model. *Journal of Physics: Conference Series*, 2596(1). <https://doi.org/10.1088/1742-6596/2596/1/012066>
- Sa'diyah, L. H., Siahaan, P., Suhendi, E., Samsudin, A., Novia, H., Komalasari, K., Umar, F. A., Dalila, A. A., Fatima, W. O., Hendriyani, D., & Purwanto, M. G. (2021). Enhancing Sundanese Students' Creative Thinking Skills Using ICARE Model on Physics Concepts: A Rasch Analysis Approach. *Journal of Physics: Conference Series*, 2098(1). <https://doi.org/10.1088/1742-6596/2098/1/012014>
- Sahida, D., & Zarvianti, E. (2019). Development of Problem Based Learning (PBL) Practicum Guide to Improve Student Creative Thinking Skills (CTS) in Basic Physics Subject. *Journal of Educational and Learning Studies*, 2(1), 39. <https://doi.org/10.32698/0492>
- Samsudin, A., Aminudin, A. H., Novia, H., Suhandi, A., Fratiwi, N. J., Yusup, M., Supriyatman, S., Masrifah, M., Adimayuda, R., Prahani, B. K., Wibowo, F. C., Faizin, M. N., & Costu, B. (2023). Identifying Javanese Students' Conceptions on Fluid Pressure with Wright Map Analysis of Rasch. *Journal of Natural Science and Integration*, 6(2), 173. <https://doi.org/10.24014/jnsi.v6i2.21822>
- Saregar, A., Cahyanti, U. N., Susilowati, N. E., Anugrah, A., & Muhammad, N. (2021). Core Learning Model: Its Effectiveness Towards Students' Creative Thinking. *International Journal of Evaluation and Research in Education*, 10(1), 35–41. <https://doi.org/10.11591/ijere.v10i1.20813>
- Soeharto, S. (2021). Development of A Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach. *Journal of Turkish Science Education*, 18(3), 351–370. <https://doi.org/10.36681/tused.2021.78>
- Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*. 173(Icei 2017), 38–42. <https://doi.org/10.2991/icei-17.2018.11>
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*.
- Sundari, P. D., & Sarkity, D. (2021). Keterampilan Berpikir Kritis Siswa SMA pada Materi Suhu dan Kalor dalam Pembelajaran Fisika. *Journal of Natural Science and Integration*, 4(2), 149. <https://doi.org/10.24014/jnsi.v4i2.11445>
- Surucu, L., & Maslakci, A. (2020). Business & Management Studies : *Business & Management Studies: An International Journal*, 8(3), 2694–2726.

- Suwistika, R., Ibrohim, I., & Susanto, H. (2024). Improving Critical Thinking and Creative Thinking Skills through POPBL Learning in High School Student. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 10(1), 115–122. <https://ejournal.umm.ac.id/index.php/jpbi/article/view/30172>
- Syukri, M., Herliana, F., Amalia, R., & Wahyuni, S. (2022). The Implementation of PBL Based on Blended Learning to Improve Students' Creative Thinking in Physics Learning. *Journal of Physics: Conference Series*, 2377(1). <https://doi.org/10.1088/1742-6596/2377/1/012085>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tennant, A., & Conaghan, P. G. (2007). The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look For in A Rasch Paper? *Arthritis Care and Research*, 57(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). Assessing Creativity: A Guide for Educators. In *Journal of Education and Learning* (Issue December). <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED505548%0Ahttp://dx.doi.org/10.1007/s41465-016-0002-3>
- Trilling, B., & Fadel, C. (2012). 21st century skills: Learning for life in our times. *John Wiley & Sons*.
- Utari, S., Manopo, J., Feranie, S., Prima, E. C., & Iswanto, B. H. (2020). Development of Standardized Online Test to Assess the Students 21st Century Skills. *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019*. <https://doi.org/10.4108/eai.12-10-2019.2296465>
- Wartono, W., Diantoro, M., & Bartlolona, J. R. (2018). Influence of Problem Based Learning Learning Model on Student Creative Thinking on Elasticity Topics A Material. *Jurnal Pendidikan Fisika Indonesia*, 14(1), 32–39. <https://doi.org/10.15294/jpfi.v14i1.10654>
- White, M., & Ronfeldt, M. (2024). Monitoring Rater Quality in Observational Systems: Issues Due to Unreliable Estimates of Rater Quality. *Educational Assessment*, 29(2), 124–146. <https://doi.org/10.1080/10627197.2024.2354311>
- Zarvianti, E., & Sahida, D. (2020). Designing Comics By Using Problem Based Learning (PBL) to Improve Student's Creative Thinking Skills. *International Journal of Social Learning (IJSL)*, 1(1), 75–88. <https://doi.org/10.47134/ijsl.v1i1.8>