EduFisika Jurnal Pendidikan Fisika

Volume 10 Number 1, April 2025, pp. 38~53
P-ISSN:2477-7935
E-ISSN: 2548-6225
DOI: 10.59052/edufisika.v10i1.42136

VALIDITY AND RELIABILITY ANALYSIS OF AN ENVIRONMENTALLY INTEGRATED SCIENCE LITERACY TEST USING RASCH MODEL

Desi Adhari[1], Santiani[2], dan Sri Fatmawati[3]
[1,2,3] Physics Education Study Program, FTIK, IAIN Palangka Raya, Indonesia
Corresponding author email: desi.adhari2111130049@iain-palangkaraya.ac.id

Article Info
Received: 28 February 2025
Accepted: 17 March 2025
Publication: 30 March 2025

Abstract :
This study aims to create an integrated assessment instrument for environmental science literacy focused on thermodynamics for high school students. The research methodology employed is Research and Development (R&D), specifically using the 4-D model. Participants in the study included 34 twelfth-grade students from Public High School 1 Palangka Raya. The validity of the instrument was evaluated using Aiken's index with input from five expert validators, while its quality was analyzed through the Rasch Model. The findings indicate that the instrument demonstrates unidimensionality, with a Raw Variance Explained by Measures value of 43.2%, confirming its effectiveness in assessing the desired aspects of science literacy. Out of the 16 items analyzed, 13 were considered fitting. The instrument also showed exceptionally high reliability, boasting a Cronbach's Alpha value of 0.89, and most items exhibited no gender bias, according to the Differential Item Functioning (DIF) test. This instrument's development is anticipated to assist students in connecting scientific concepts to real-world scenarios, enhance their science literacy skills, and raise awareness of the importance of environmental conservation, particularly regarding peatland preservation in Palangka Raya.

Keywords: Environment, Rasch Model, Science Literacy

INTRODUCTION

21st-century skills have become a primary focus in today's global education system (Kain et al., 2024; Nisrina et al., 2020). Education in this era aims to equip students with competencies that enable them to adapt to the dynamic changes of the times (Roshid & Haider, 2024; Sutrisna, 2021). The World Economic Forum (WEF) has identified 16 essential skills for the 21st century, with scientific literacy being one of them (Yusmar & Fadilah, 2023). Scientific literacy plays a crucial role in shaping scientific thinking patterns and supporting evidence-based decision-making in everyday life (OECD, 2024).

In Indonesia, scientific literacy has been integrated into the curriculum since the implementation of the 2006 School-Based Curriculum, the 2013 Curriculum (K-13), and the Merdeka Curriculum (Kurmer) (Yusmar & Fadilah, 2023). Conceptually, scientific literacy encompasses the ability to utilize scientific knowledge to identify problems, acquire new information, and explain phenomena using an evidence-based approach (OECD, 2024). Scientific literacy, when combined with an understanding of science and technology, plays a crucial role in shaping the decision-making patterns of future generations (Ramli et al., 2022). In line with this, data literacy also aims to equip students with critical and reflective thinking skills regarding information and its impact on social and political life (Fagerlund et al., 2025).

The scientific literacy achievement of Indonesian students is at an alarming level, as indicated by the results of the Programme for International Student Assessment (PISA). The 2018 PISA results reported that Indonesian students scored 396 in scientific literacy, showing a decline compared to the 2015 PISA results, which ranged between 382 and 403 (Kemendikbud, 2019). The latest PISA 2022 report further revealed that students' scientific literacy scores dropped to 383, with the majority of students classified at Level 1a—the second-lowest level in the OECD classification (OECD, 2024). The average scientific literacy score among OECD countries is between 483 and 488, significantly higher than that of Indonesian students. Moreover, the 2022 PISA survey highlighted that not a single Indonesian student achieved Level 5 or 6 in scientific literacy, indicating a lack of ability to apply scientific knowledge creatively and independently across various contexts (OECD, 2023).

The low level of scientific literacy in Indonesia is attributed to several factors, including suboptimal curriculum implementation, limited variation in teaching methods, inadequate educational facilities and resources, and the lack of valid and reliable assessment instruments to measure students' scientific literacy (Fuadi et al., 2020; Yusmar & Fadilah, 2023). The attainment of scientific literacy in schools is influenced by various aspects of the educational environment, including student, teacher, and school management factors. The interaction between students' attitudes, aspirations, and mental well-being, along with the quality of teaching and school policies, forms a system that determines the effectiveness of science learning (Ding, 2022; Holzberger et al., 2020). One of the key weaknesses in current assessment practices is the dominance of test items that focus solely on conceptual aspects without integrating scientific understanding into real-world contexts, particularly environmental issues.

One strategy to enhance scientific literacy is to provide more contextual and real-world-based learning experiences. Research findings indicate that student engagement in experiential activities, such as internships or scientific projects, enhances their knowledge, attitudes, and 21st-century skills. Participation in such activities also motivates students to pursue higher education, particularly in science-related fields (Saraiva et al., 2025). Therefore, the development of a science literacy evaluation tool that not only measures conceptual understanding but also the skills of applying science in everyday life is very important (Rosana et al., 2020). According to the PISA framework, scientific literacy consists of four main dimensions: scientific competence, scientific knowledge, scientific context, and scientific attitudes. The competence dimension serves as a fundamental aspect in measuring scientific literacy, as it reflects students' ability to apply science in solving real-world problems (OECD, 2023).

One branch of science that is closely related to environmental issues is thermodynamics, which examines the relationship between energy, work, and entropy in various physical systems (Tudor, 2023). Understanding thermodynamic concepts such as the law of energy conservation, energy efficiency, and energy degradation is crucial for students in analyzing environmental problems (De Hemptinne et al., 2022), including climate change, air pollution, and peatland fires. Therefore, integrating scientific literacy with environmental aspects in thermodynamics learning can serve as an effective strategy to enhance students' understanding of scientific concepts while fostering their ecological awareness.

The results of observations conducted at Public High School 1 Palangka Raya indicate that students' scientific literacy levels are still low, with only 42% of students answering correctly. Less than 50% of students were able to answer the given scientific literacy questions correctly. These findings were obtained from one 12th-grade class consisting of 32 students. Based on the observations, the evaluation instruments used by the students were more oriented toward measuring conceptual understanding without emphasizing the application of scientific knowledge in real-world contexts, such as environmental issues, particularly peatlands in Palangka Raya.

This study develops a scientific literacy assessment tool integrated with environmental issues in thermodynamics using the Rasch Model approach. Unlike previous studies, such as (Rosana et al., 2020), which developed a PISA-based scientific literacy instrument for middle school students, and (Wilsa et al., 2023), which examined the validity and reliability of a scientific literacy instrument in cell biology, this research specifically integrates environmental aspects, particularly peatland ecology, into the assessment of students' scientific literacy. The analysis using the Rasch Model provides deeper insights into the validity, reliability, and potential bias of the instrument, enabling a more precise mapping of students' abilities (Sumintono & Widhiarso, 2015).

This study fills the gap by developing a scientific literacy assessment tool that not only measures the understanding of thermodynamics concepts but also integrates environmental issues, particularly peatland ecosystems. The urgency of this research is supported by the increasing need for more contextual and real-world-based assessments, in line with 21st-century learning principles and the implementation of the *Merdeka* Curriculum. This instrument can assist educators in assessing and improving students' scientific literacy in a context more relevant to everyday life. Additionally, this study supports the implementation of the *Kurikulum Merdeka*, which emphasizes exploration-based learning and real-world problem-solving (Kemendikbud, 2019). Schools can also utilize the findings of this study to design more contextual assessments, integrate scientific literacy with environmental issues, and assist teachers in developing more effective teaching strategies to enhance students' environmental awareness.

Based on the discussion above, this study aims to develop an environmental-integrated scientific literacy assessment instrument for thermodynamics material in high school students. The development of this instrument is expected to help students connect scientific concepts with real-life situations, enhance their scientific literacy, and foster awareness of the importance of environmental conservation.

RESEARCH METHOD

Research Design

This study employs the Research and Development (R&D) method, a research approach aimed at developing a specific product and evaluating its effectiveness (Sugiyono, 2012). The research design follows the 4-D model (Four-D Model), which consists of four stages: define, design, develop, and disseminate (Thiagarajan & Semmel, 1974).

The define stage includes several steps: preliminary-final analysis, student analysis, and final task analysis. The second stage, design, involves the development of a scientific literacy-based assessment instrument integrated with environmental aspects. The instrument is structured in the form of an extended response question aligned with the scientific literacy dimensions of PISA 2018, particularly focusing on the competency dimension, as it effectively represents students' scientific literacy (Rosana et al., 2020). The development stage aims to examine the content validity and empirical validity of the assessment instrument through Rasch Model analysis. The final stage, dissemination, is conducted through the publication of scientific articles.

Research Subjects

This study involved a population of 12th-grade students from classes 1 to 6 in the second semester of the 2024/2025 academic year at Public High School 1 Palangka Raya who were still actively attending lessons. The research sample consisted of class XII-2 at Public High School 1 Palangka Raya in the 2024/2025 academic year, with a total of 34 students, comprising 22 female students and 12 male students.

Data Collection Techniques

Questionnaire

A questionnaire is a data collection tool consisting of a series of written questions given to respondents to obtain written answers (Supriadi et al., 2020). In this study, the questionnaire was used to obtain validation from experts and practitioners regarding the developed instrument.

Test

A test is a method or procedure used in measurement and assessment in the field of education. It typically involves assigning tasks or a series of questions that must be answered by students, allowing the obtained data to represent their behavior or competencies (Sudijono, 2011). In this study, the instrument used was a scientific literacy test integrated with environmental contexts in the form of essay questions. This instrument was designed to analyze the validity and reliability of the test items using the Rasch Model.

Data Analysis Techniques

Content Validity Analysis

The content validity of the instrument in this study was assessed using Aiken's V index, involving five expert validators. The validation process evaluated the instrument based on three key aspects: content, scientific literacy, and language clarity. The content aspect ensured the alignment of the questions with thermodynamics concepts and their appropriateness for the educational level. The scientific literacy aspect examined the extent to which the instrument effectively measured students' scientific literacy competencies in accordance with the PISA framework. The language clarity aspect ensured that the questions were structured clearly and unambiguously. The Aiken's V formula used to calculate the validity index is presented in Equation 1 (Aiken, 1985).

$$V = \frac{\Sigma s}{n\,(c-1)} \quad (1)$$

The validity assessment was conducted by comparing V_{Value} with V_{table}. Suppose $V_{Value} > V_{table}$ the instrument is categorized as valid. The critical V_{table} value for five raters at a 5% significance level is 0.80. Therefore, if $V_{Value} > 0.80$, the instrument is considered valid; otherwise, if $V_{Value} < V_{table}$, it is deemed invalid (Aiken, 1985). The results of this validation process serve as the basis for revising the instrument before its field trial implementation.

Instrument Quality Analysis Using the Rasch Model

Data analysis was conducted using the Rasch Model with Winsteps software version 4.8.2.0, enabling the evaluation of various psychometric measurement aspects. Several key aspects were analyzed, including a unidimensionality test to ensure that the instrument measures only a single construct, item fit analysis to assess how well each item aligns with the Rasch model, and item difficulty analysis to determine the difficulty level of each question. Additionally, instrument reliability was examined to ensure measurement consistency, while a variable map was utilized to visualize the relationship between item difficulty levels and respondents' abilities. The final analysis, Differential Item Functioning (DIF), was conducted to identify potential bias in items based on specific respondent characteristics, such as gender or other demographic groups. DIF analysis aims to ensure that no item provides undue advantages or disadvantages to any group, thereby maintaining fairness and aligning with the objectives of the scientific literacy evaluation (Boone et al., 2014; Sumintono & Widhiarso, 2014).

RESULTS AND DISCUSSION

This study resulted in the development of an environmental-based scientific literacy assessment instrument for high school students in the thermodynamics domain. The instrument was developed through the stages of the 4-D model, underwent trial testing, and was refined based on feedback from expert reviewers.

Findings from the definition stage indicate that teachers in schools had never administered a scientific literacy-based assessment linked to environmental issues. However, observations revealed that schools have attempted to introduce environmental aspects through the Projek Penguatan Profil Pelajar Pancasila (P5) initiative. The lack of integration of environmental issues in learning assessments represents a gap that needs to be addressed. Observations of students further revealed that their level of scientific literacy was significantly low. This finding aligns with PISA reports, which indicate that Indonesian students' scientific literacy performance remains relatively low compared to other countries (OECD, 2019). A more contextual and environmentally integrated assessment is crucial for enhancing students' scientific literacy, as suggested by (Bybee, 2013) through the Science, Technology, and Society (STS) learning model, which emphasizes the connection between science and real-world problems.

The design stage of the environmental-integrated scientific literacy evaluation tool in the thermodynamics domain for high school students was developed based on one of the dimensions introduced by the PISA 2018 competency dimension, which consists of three key indicators (OECD, 2019). The developed instrument blueprint is presented in Table 1.

Table 1. Science Literacy Instrument Grid Integrated with Environmental Issues in Thermodynamics

Competency Dimension	Indicator	Material
Explaining Scientific Phenomena	Recall and apply appropriate scientific knowledge.	First Law of Thermodynamics (C1 & C3)
	Identify, utilize, and generate clear models and representations.	Isothermal Compression Process & Carnot Engine Cycle (C4 & C5)
	Explain the potential implications of scientific knowledge for society.	Earth's Heat Energy and Peatland & Renewable Energy (C5 & C4)
Designing and Evaluating Scientific Investigations	Formulate and explore scientific questions systematically.	Solar Water Heater Experiment & Heat Engine Efficiency (C3 & C4)
	Evaluate how scientific questions are explored.	Heat Energy Conversion in Closed Systems & Energy Efficiency in Thermodynamic Systems (C5)
	Describe and evaluate various methods used by scientists to determine the validity and objectivity of data and conclusions.	Measuring Peatland Temperature & Validating Heat Energy Data (C4)
Transforming Scientific Data and Evidence	Convert data from one representation to another.	PV Graph & Temperature Change Graph of Burned Peatland (C3 & C4)
	Analyze and interpret data to draw accurate conclusions.	Biomass Energy Efficiency on Peatland & Heat Analysis of Burned Peatland (C5 & C4)

In Table 1, the three competency dimension indicators are further elaborated based on previous research conducted (Setiawan, 2019). his assessment instrument is designed by considering cognitive levels ranging from C3 (Applying) to C5 (Evaluating). The selection of these cognitive levels aims to assess students' ability to apply thermodynamic concepts in various contexts, including environmental issues. The assessment instrument is structured not only to measure students' conceptual understanding but also to evaluate their ability to analyze and assess scientific phenomena in real-world situations.

The development phase was carried out to ensure that the developed product met high-quality standards. Before the field trial, students were provided with a leaflet containing insights about their surrounding environment, particularly the peatland ecosystem, as an effort to introduce them to environmental issues relevant to the science literacy assessment. With this leaflet, students were expected to gain a better understanding of the connection between thermodynamic concepts and environmental problems, enabling them to answer the questions more contextually.

Analysis of Content Validity

The content validity evaluation of the instrument was conducted by five experts using Aiken’s Index to assess the alignment of the developed test items with aspects of content, scientific literacy, and language. This assessment covered 16 test items, with the results presented in Table 2.

Table 2. Expert Validation Analysis Results Using Aiken's Index

Item	V_{Value}	Criteria	Item	V_{Value}	Criteria
1	0.9	Valid	9	0.93	Valid
2	0.91	Valid	10	0.88	Valid
3	0.87	Valid	11	0.91	Valid
4	0.9	Valid	12	0.94	Valid
5	0.92	Valid	13	0.91	Valid
6	0.93	Valid	14	0.91	Valid
7	0.89	Valid	15	0.88	Valid
8	0.89	Valid	16	0.88	Valid

Based on Table 2, the calculated V_{Value} for each test item ranges from 0.87 to 0.94. According to Aiken's validity criteria (Aiken, 1985). An instrument is considered valid if V_{Value} > 0,80. Since all test items have a V_{Value} greater than 0.80, they are deemed valid, with some revisions suggested by experts. The content validity results indicate that the test items have met the substantive aspects of the material in accordance with thermodynamics concepts and the appropriate educational level. Additionally, the items have a clear structure, ensuring that no ambiguity arises in students' comprehension.

The validation results demonstrate that the evaluation instrument has met the required content validity standards, making it suitable for use in research. These findings align with the principles outlined by Arikunto, who stated that a valid instrument must accurately measure the targeted aspects and align with the intended assessment objectives (Arikunto, 2019).

When compared to previous studies, the content validity of this instrument shows consistency with the research conducted by (Wilsa et al., 2023), which reported an average Aiken's V score of 0.90, as well as the study by (Sheptian et al., 2024) which found that its instrument had strong content validity with an Aiken's V score exceeding 0.80. The developed instrument exhibits a comparable level of validity to prior studies. However, this instrument has the added advantage of integrating scientific literacy and environmental aspects into thermodynamics material. This area has been relatively unexplored in previous research, as highlighted in the literature analysis by (Adhari et al., 2024).

Analisis Unidimensionality Instrumen

The validity analysis using the Rasch Model includes an assessment of the instrument's unidimensionality. This concept refers to the ability of the instrument to consistently and accurately measure a single targeted construct (Sumintono & Widhiarso, 2015). Unidimensionality ensures that each item within the instrument effectively evaluates the intended aspect of the study, making the assessment results reliable. The results of the unidimensionality analysis are presented in Table 3.

Table 3. Unidimensionality analysis

Keterangan	**Empirical**			**Modeled**
Total raw variance in observations	28.18	100%		100.0%
Raw variance explained by measures	12.18	43.2%		44.0%
Raw variance explained by persons	1.74	6.2%		6.3%
Raw Variance explained by items	10.44	37.0%		37.7%
Raw unexplained variance (total)	16.0	56.8%	100.0%	56.0%
Unexplned variance in 1st contrast	2.98	10.6%	18.7%	
Unexplned variance in 2nd contrast	2.33	8.3%	14.6%	
Unexplned variance in 3rd contrast	1.71	6.1%	10.7%	
Unexplned variance in 4th contrast	1.67	5.9%	10.4%	
Unexplned variance in 5th contrast	1.54	5.5%	9.6%	

To determine whether the instrument meets the unidimensionality criteria, two main conditions must be satisfied. First, the Raw Variance Explained by Measure must be at least 20%. If the value falls within the 20%–40% range, the instrument is categorized as adequate; a value between 40%–60% indicates good quality, while a score above 60% is considered excellent (Sumintono & Widhiarso, 2014). Second, the Unexplained Variance in the first contrast should be less than 15%, indicating that external factors beyond the primary construct do not significantly influence the measurement results. If both conditions are met, the instrument is considered to have strong validity in assessing the designated aspect (Sumintono & Widhiarso, 2014).

Based on the analysis using the Rasch Model, as presented in Table 3, the Raw Variance Explained by Measures was found to be 43.2%. An instrument with a value exceeding 40% is considered to have a strong ability to explain the variance expected by the model. According to (Sumintono & Widhiarso, 2015), this value falls within the "good" category, as it lies within the 40%–60% range. This indicates that the instrument is valid and appropriate for measuring students' scientific literacy competencies in thermodynamics.

Additionally, the Unexplained Variance in the first contrast was 10.6%, while the unexplained variance in the second to fifth contrasts was 8.3%, 6.1%, 5.9%, and 5.5%, respectively. Since all these values are below the 15% threshold, it suggests that no dominant factors outside the main construct being measured influenced the results. Therefore, the instrument meets the Unidimensionality criteria. In other words, no evidence was found indicating the presence of other dimensions that could compromise the instrument's validity (Sumintono & Widhiarso, 2015).

Item Fit Analysis

The item fit level refers to the extent to which an item functions as expected within a measurement model. Well-fitting items ensure that students clearly understand the questions, allowing their responses to accurately reflect the intended ability or attitude being measured rather than being influenced by confusion over the question's meaning (Sumintono & Widhiarso, 2015). An item is considered fit if it meets the following three key parameters: 0.5 < MNSQ < 1.5, -2.0 < ZSTD < +2.0, dan 0.4 < Pt Mean Corr < 0.85 (Sumintono & Widhiarso, 2014). The results of the item fit analysis are presented in Figure 1.

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	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
2	99	34	-.10	.10	2.38	4.29	2.47	4.41	A .42	.62	5.9	21.3	S2
4	129	34	-.40	.10	2.26	4.10	2.14	3.80	B .59	.65	8.8	21.1	S4
13	131	34	-.42	.10	1.01	.10	1.09	.43	C .46	.65	14.7	21.1	S13
10	98	34	-.09	.10	1.02	.15	1.02	.17	D .76	.62	11.8	21.3	S10
14	63	34	.32	.12	.97	-.03	1.00	.10	E .63	.56	11.8	24.4	S14
8	93	34	-.04	.10	.95	-.11	.99	.02	F .54	.62	23.5	22.8	S8
3	93	34	-.04	.10	.92	-.28	.96	-.10	G .65	.62	23.5	22.8	S3
16	66	34	.28	.11	.96	-.09	.91	-.26	H .61	.56	17.6	24.4	S16
15	88	34	.02	.10	.95	-.11	.94	-.17	h .69	.61	11.8	22.9	S15
5	73	34	.19	.11	.73	-1.13	.77	-.84	g .58	.58	41.2	22.9	S5
12	69	34	.24	.11	.77	-.92	.76	-.87	f .64	.57	26.5	24.2	S12
9	62	34	.33	.12	.66	-1.41	.60	-1.57	e .66	.55	32.4	24.5	S9
1	80	34	.11	.11	.63	-1.65	.63	-1.58	d .56	.59	29.4	23.0	S1
7	112	34	-.23	.10	.63	-1.75	.62	-1.77	c .70	.64	29.4	21.6	S7
11	83	34	.07	.11	.53	-2.25	.60	-1.78	b .69	.60	38.2	23.0	S11
6	114	34	-.25	.10	.55	-2.24	.57	-2.10	a .80	.64	26.5	21.7	S6
MEAN	90.8	34.0	.00	.11	.99	-.2	1.00	-.1			22.1	22.7	
P.SD	21.5	.0	.23	.01	.53	1.8	.52	1.8			10.3	1.2	

Figure 1. Item Fit Analysis

Based on the analysis results using the Rasch Model, it was found that 13 out of 16 items were deemed fit as they met the three main parameters, while three items (S2, S4, and S6) were classified as misfits. This indicates that the majority of the questions align with the expected measurement model for assessing students' abilities (Danielis et al., 2025; Sumintono & Widhiarso, 2015). Items S2 and S4 had MNSQ and ZSTD values significantly exceeding the established thresholds, suggesting that these questions might be too difficult or contain unclear sentence structures, potentially confusing students (Boone et al., 2014). Additionally, item S6 had a ZSTD value outside the acceptable range (-2.10), indicating that the question was too easy, making it less effective in distinguishing students' ability levels (Bond & Fox, 2015). Therefore, items S2, S4, and S6 need revision by refining the wording of the questions or conducting a readability test with students to ensure better comprehension.

The validity analysis of the test items in this study revealed that 13 out of 16 questions were classified as fit, while three items were misfit and required revision. A previous study conducted by (Rahmati et al., 2024) found that out of 20 test items, only 10 met all three fit criteria. In comparison, the instrument in this study demonstrated a higher item validity level, as a greater percentage of items met the fit criteria (81.25% compared to 50%). This difference may be attributed to instrument design aspects, the alignment of questions with the measured competencies, or the characteristics of the sample used in the trial. Nevertheless, the findings from both studies emphasize the importance of validating test items using the Rasch Model to ensure that the instrument accurately measures the targeted competencies.

Item Difficulty Level Analysis

The difficulty level of test items in the Rasch Model is categorized based on the Measure logit and the Standard Deviation (SD) logit of the items. It is divided into four categories: very difficult (greater than +1 SD), difficult (0.00 logit + 1 SD), easy (0.00 logit – 1 SD), and very easy (less than -1 SD) (Sumintono & Widhiarso, 2015). The results of the item difficulty level test are presented in Figure 2.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	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
9	62	34	.33	.12	.66	-1.41	.60	-1.57	.66	.55	32.4	24.5	S9
14	63	34	.32	.12	.97	-.03	1.00	.10	.63	.56	11.8	24.4	S14
16	66	34	.28	.11	.96	-.09	.91	-.26	.61	.56	17.6	24.4	S16
12	69	34	.24	.11	.77	-.92	.76	-.87	.64	.57	26.5	24.2	S12
5	73	34	.19	.11	.73	-1.13	.77	-.84	.58	.58	41.2	22.9	S5
1	80	34	.11	.11	.63	-1.65	.63	-1.58	.56	.59	29.4	23.0	S1
11	83	34	.07	.11	.53	-2.25	.60	-1.78	.69	.60	38.2	23.0	S11
15	88	34	.02	.10	.95	-.11	.94	-.17	.69	.61	11.8	22.9	S15
3	93	34	-.04	.10	.92	-.28	.96	-.10	.65	.62	23.5	22.8	S3
8	93	34	-.04	.10	.95	-.11	.99	.02	.54	.62	23.5	22.8	S8
10	98	34	-.09	.10	1.02	.15	1.02	.17	.76	.62	11.8	21.3	S10
2	99	34	-.10	.10	2.38	4.29	2.47	4.41	.42	.62	5.9	21.3	S2
7	112	34	-.23	.10	.63	-1.75	.62	-1.77	.70	.64	29.4	21.6	S7
6	114	34	-.25	.10	.55	-2.24	.57	-2.10	.80	.64	26.5	21.7	S6
4	129	34	-.40	.10	2.26	4.10	2.14	3.80	.59	.65	8.8	21.1	S4
13	131	34	-.42	.10	1.01	.10	1.09	.43	.46	.65	14.7	21.1	S13
MEAN	90.8	34.0	.00	.11	.99	-.2	1.00	-.1			22.1	22.7	
P.SD	21.5	.0	.23	.01	.53	1.8	.52	1.8			10.3	1.2	

Figure 2. Item Difficulty Level Analysis

Figure 2, Item Measure Order Output, analyzes the difficulty level of instrument items. The data indicates that the Standard Deviation (SD) value for the instrument items is 0.23. By combining this value with the logit mean, the difficulty level of the test items can be classified, as shown in Table 4.

Table 4. Difficulty Level Categories

Item Difficulty	Category
>0.23	Very Difficult
0.00 – 0.23	Difficult
-0.23 – 0.00	Easy
< -0.23	Very Easy

Based on the analysis results using the Rasch Model, as shown in Figure 2, the difficulty level of the test items in the developed scientific literacy evaluation instrument demonstrates a fairly balanced distribution. Items categorized as very difficult are found in questions number 9, 14, 16, and 12, while difficult questions include numbers 5, 1, 11, and 15. Conversely, questions classified as easy include numbers 3, 8, 10, 2, and 7, while three other questions are categorized as very easy, namely questions number 6, 4, and 13. The percentage distribution of item difficulty levels is presented in Table 5.

Table 5. Percentage of Instrument Difficulty Level

Item	Category	Persentase
9, 14, 16, dan 12	Very Difficult	25%
5, 1, 11, dan 15	Difficult	25%
3, 8, 10, 2, dan 7	Easy	31%
6, 4, dan 13	Very Easy	19%

Compared to previous research conducted by Rahmati et al. (2024), the distribution of item difficulty levels in this study exhibits a more balanced pattern in terms of item spread across difficulty categories. In Ulvi Rahmati's study, the majority of the test items fell into the easy (40%) and difficult (35%) categories, while the very easy category accounted for only 5%, and the very difficult category comprised 20%.

In contrast, this study demonstrates a more even distribution of difficulty levels, with very difficult and difficult items, each representing 25%, easy items 31%, and very easy items 19%. A well-distributed difficulty level in an instrument should encompass a range of student abilities to provide more comprehensive diagnostic information (Mardapi, 2017; Sumintono & Widhiarso, 2015). The developed instrument has met the principle of difficulty level balance in psychometric measurement. However, further analysis of items classified as very easy is necessary to ensure that the instrument remains challenging yet appropriate for students' cognitive levels.

Instrument Reliability Analysis

The instrument reliability test aims to ensure that the instrument used can generate consistent data in repeated measurements (Nurlatifah et al., 2023). In the analysis using the Rasch Model, reliability is assessed based on two main indicators, namely Item Reliability and Person Reliability, which respectively represent the consistency of test items and the stability of student responses. The criteria for determining the reliability values for both indicators are presented in Table 6.

Table 6. Instrument Reliability Criteria

Interval	Category
0.0 – 0.19	Very Low
0.20 – 0.39	Low
0.40 – 0.59	Moderate
0.60 – 0.79	High
0.80 – 1.00	Very High

(Arikunto, 2010; Sugiyono, 2014)

The grouping of persons and items can be analyzed based on the separation value. The higher the separation value, the better the instrument's quality in distinguishing respondent groups and test items overall (Sumintono & Widhiarso, 2014). The results of the reliability test for the scientific literacy instrument are presented in Figure 3.

```
     SUMMARY OF 34 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      42.7      16.0        -.59     .16      1.04    .02   1.00   -.05 |
|  SEM       3.6        .0         .09     .01       .09    .23    .08    .23 |
| P.SD      20.6        .0         .49     .04       .50   1.34    .48   1.29 |
| S.SD      20.9        .0         .50     .04       .51   1.36    .49   1.31 |
| MAX.      85.0      16.0         .27     .30      2.64   3.52   2.60   3.44 |
| MIN.       8.0      16.0       -1.75     .13       .31  -2.74    .31  -2.75 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .18 TRUE SD     .46  SEPARATION  2.54  Person RELIABILITY  .87 |
|MODEL RMSE    .16 TRUE SD     .46  SEPARATION  2.82  Person RELIABILITY  .89 |
| S.E. OF Person MEAN = .09                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89  SEM = 6.81
STANDARDIZED (50 ITEM) RELIABILITY = .96

     SUMMARY OF 16 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      90.8      34.0         .00     .11       .99   -.21   1.00   -.13 |
|  SEM       5.5        .0         .06     .00       .14    .48    .14    .46 |
| P.SD      21.5        .0         .23     .01       .53   1.85    .52   1.79 |
| S.SD      22.2        .0         .24     .01       .54   1.91    .54   1.85 |
| MAX.     131.0      34.0         .33     .12      2.38   4.29   2.47   4.41 |
| MIN.      62.0      34.0        -.42     .10       .53  -2.25    .57  -2.10 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .11 TRUE SD     .20  SEPARATION  1.80  Item   RELIABILITY  .76 |
|MODEL RMSE    .11 TRUE SD     .21  SEPARATION  1.96  Item   RELIABILITY  .79 |
| S.E. OF Item MEAN = .06                                                     |
-------------------------------------------------------------------------------
```

Figure 3. Instrument Reliability Test

The results of the instrument reliability analysis using the Rasch Model, as shown in Figure 3, indicate that the obtained Cronbach's Alpha value is 0.89. This value suggests that the overall interaction between respondents and test items falls into the very high category. The high Cronbach's Alpha value confirms that the instrument possesses good internal consistency and can be reliably used to assess students' abilities (Bond & Fox, 2015).

The person reliability value of 0.87 indicates that students' responses to this instrument demonstrate high consistency. From the perspective of item quality, the item reliability value of 0.76 suggests that the test items are of good quality. A previous study conducted by (Krisanda and Harjito, 2021) reported an item reliability value of 0.78 and a person reliability value of 0.41. The reliability analysis in this study shows higher values.

A high person reliability value indicates that respondents answered consistently, allowing the instrument to accurately differentiate students' ability levels (Boone et al., 2014). Meanwhile, a high item reliability value signifies that the instrument contains high-quality test items that effectively measure the targeted skills or competencies (Sumintono & Widhiarso, 2015). The relatively comparable item reliability values suggest that the quality of test items in both studies falls within the good category. These differences may be attributed to variations in instrument design, sample population, or the context of the tested material.

Analysis of Variable Map

The Variable Map illustrates the distribution of persons (respondents' abilities) and items (item difficulty levels). The scale used is the logit scale, where positive values indicate higher ability or difficulty levels, while negative values represent lower levels (Boone et al., 2014).

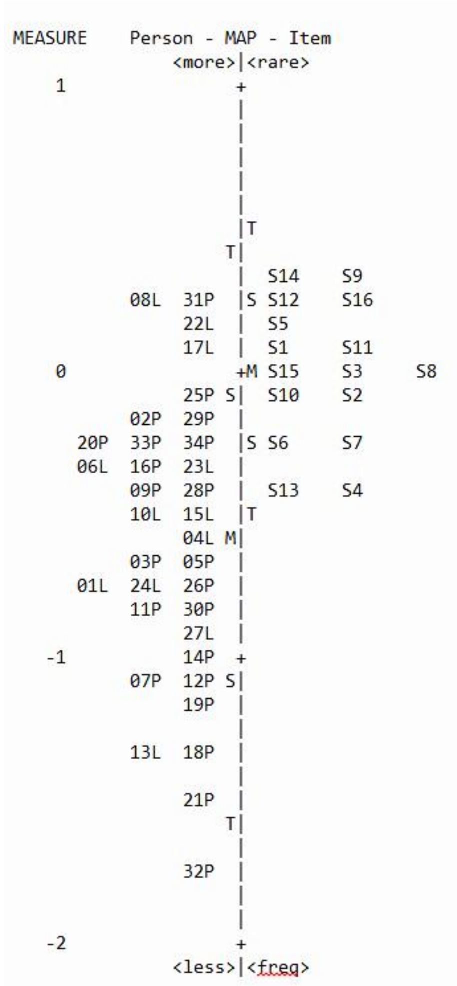

Figure 4. Variabel Map

Based on the analysis in Figure 4, the distribution of respondents' abilities indicates that most participants fall within the range of -1 to +1 logit. Respondents with the highest abilities are around +1 logit, including participants with codes 08L, 31P, 22L, and 17L, while those with lower abilities (around -1 to -2 logit) are 13L, 18P, 21P, and 32P. This suggests that the majority of participants possess moderate abilities, with only a few falling into the categories of very high or very low ability.

The test items range in difficulty from the most difficult to the easiest. The most difficult items are S14 and S9, which are located around +1 logit, while items with moderate difficulty (around 0 logit) include S15, S10, and S2. The easiest items, located around -1 logit, are S13 and S4. This distribution demonstrates that the developed instrument has a well-balanced variation in difficulty levels.

Differential Item Function (DIF) Analysis

Differential Item Function (DIF) is an essential technique in survey and test data analysis and is also necessary when analyzing data using the Rasch model (Boone et al., 2014). DIF assesses whether an item is biased or favors a particular characteristic (Soeharto et al., 2024; Sumintono & Widhiarso, 2014). An item is considered biased if it has a PROB value of less than 0.05 (Boone et al., 2014).

Based on the analysis in Figure 5, most items have a PROB value greater than 0.05, indicating that these items do not show significant differences in difficulty or ease of completion between male (L) and female (P) respondent groups. The majority of the test items in this instrument can be considered fair and do not exhibit bias toward either gender group. However, one item has a PROB value of 0.0485, suggesting a potential bias that requires further analysis to determine its cause. The Differential Item Function (DIF) analysis in graphical form is presented in Figure 6.

Person CLASSES	SUMMARY DIF CHI-SQUARED	D.F.	PROB.		BETWEEN-CLASS/GROUP UNWTD MNSQ	ZSTD	Item Number	Name
2	1.8080	1	.1787		1.9567	1.00	1	S1
2	.2529	1	.6151		.2639	-.29	2	S2
2	2.2022	1	.1378		2.4062	1.19	3	S3
2	1.2525	1	.2631		1.3416	.69	4	S4
2	1.3787	1	.2403		1.4808	.77	5	S5
2	.0174	1	.8951		.0271	-1.01	6	S6
2	.0000	1	1.0000		.0085	-1.22	7	S7
2	1.3850	1	.2393		1.4839	.77	8	S8
2	1.0275	1	.3107		1.0918	.53	9	S9
2	.0590	1	.8082		.0616	-.81	10	S10
2	.0000	1	1.0000		.0137	-1.14	11	S11
2	.0000	1	1.0000		.0010	-1.44	12	S12
2	3.8940	1	.0485	b	4.4413	1.84	13	S13
2	2.6551	1	.1032		2.9404	1.39	14	S14
2	3.0949	1	.0785		3.4616	1.56	15	S15
2	1.3081	1	.2527		1.4028	.72	16	S16

Figure 5. Differential Item Function (DIF) Analysis

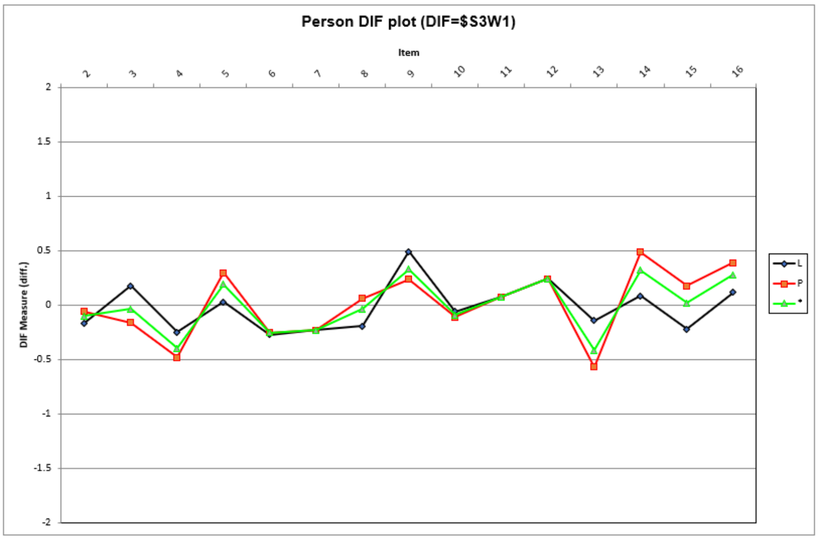

Figure 6. Differential Item Function (DIF) Analysis Graph by Gender

Based on the graph, question number 13 is identified as biased, with a PROB value of 0.0485 and a DIF Measure value for female students (-0.57) lower than that for male students (-0.14), indicating that this question is easier for female students to answer (Krisanda & Harjito, 2021). These findings are consistent with previous studies that have identified gender bias in certain test items. For example, (Wahyuni, 2022) detected gender bias through DIF analysis on school exam test items in Yogyakarta.

This study uses the Rasch Model to evaluate the validity and reliability of a science literacy assessment instrument integrated with environmental issues in thermodynamics for high school students. Expert validation results indicate that all test items achieved V values ranging from 0.87 to 0.94, exceeding Aiken's validity threshold (>0.80). This confirms that the instrument meets the substantive content criteria, aligning with thermodynamics concepts and the targeted educational level.

The distribution of item difficulty levels demonstrates a well-balanced variation, with test items spanning easy to difficult categories. Validity analysis using the Rasch Model further confirms that the instrument exhibits strong construct validity, as evidenced by item fit analysis and model alignment. The instrument's reliability is also supported by high person reliability and item reliability values, indicating both consistent respondent performance and high-quality test items.

The Raw Variance Explained by the Measure value of 43.2% suggests that the instrument effectively captures the intended dimensions of scientific literacy. The analysis of respondent ability distribution reveals that most students fall within the moderate ability range, with only a few classified as very high or very low. Furthermore, Differential Item Functioning (DIF) analysis indicates that nearly all test items are free from gender bias, except one item that requires further refinement.

The findings of this study reinforce those of (Wilsa et al., 2023), who developed a science literacy instrument based on the Rasch Model in the field of biology, demonstrating high validity and reliability. The key advantage of this study lies in integrating environmental aspects into science literacy assessment, particularly concerning peatland ecosystems, which were not covered in Wilsa et al.'s study. Meanwhile, (Rosana et al., 2020) developed a PISA-based science literacy instrument, but its scope was limited to middle school students and did not explicitly integrate environmental issues. Thus, this study addresses the gap in previous research by providing a more contextually relevant instrument for high school students to understand thermodynamics concepts through an environmental perspective.

The study conducted by (Rahmati et al., 2024) is also relevant, as it similarly developed a science literacy instrument using the Rasch Model. However, there are significant differences in the quality of the resulting instruments. Analysis results indicate that out of 20 test items examined in Rahmati et al.'s study, only 10 items (50%) met the fit criteria. In contrast, the present study demonstrated superior performance, with 13 out of 16 items (81.25%) classified as fit. This percentage difference is likely attributed to the greater contextual relevance of environmental integration in this study, which aligns more closely with students' understanding and a more targeted instrument design that focuses on specific science literacy competencies.

Compared to the study by (Krisanda & Harjito, 2021), which reported an item reliability score of 0.78 and a person reliability score of 0.41, this study demonstrates an advantage with a higher person reliability score and an almost equivalent item reliability score. According to (Boone et al., 2014), high person reliability indicates that respondents answered consistently, allowing the instrument to differentiate students' ability levels accurately. This advantage suggests integrating environmental contexts into science literacy assessments enhances students' response consistency. Meanwhile, the high item reliability score aligns with the findings of (Sumintono & Widhiarso, 2015), which emphasize that a well-designed instrument ensures high-quality test items in measuring the intended competencies.

Novelty

Using the Rasch Model approach, this study develops a scientific literacy assessment tool integrated with environmental issues in thermodynamics. Unlike previous studies, such as (Rosana et al., 2020), which developed a PISA-based scientific literacy instrument for middle school students, and (Wilsa et al., 2023), which examined the validity and reliability of a scientific literacy instrument on cell biology, this research specifically integrates environmental aspects, particularly peatland ecology, into the measurement of students' scientific literacy. The analysis using the Rasch Model provides a deeper insight into the instrument's validity, reliability, and potential bias, enabling a more accurate mapping of students' abilities (Sumintono & Widhiarso, 2015).

Implications

This instrument can assist educators in assessing and enhancing students' scientific literacy in a context more relevant to everyday life. Additionally, this study supports the implementation of the *Kurikulum Merdeka*, which emphasizes exploration-based learning and real-world problem-solving (Kemendikbud, 2019). Schools can also utilize the findings of this study to design more contextual assessments, integrate scientific literacy with environmental issues, and help teachers develop more applicable teaching strategies to enhance students' environmental awareness.

Limitations

This study presents several limitations. Firstly, its scope is restricted to a single school and involves a sample of just 34 students, which constrains the generalizability of the findings. Secondly, while the instrument has undergone testing using the Rasch Model, this research has not assessed its long-term impact on enhancing students' scientific literacy. Additionally, the environmental issues tackled are limited to the peatland ecosystem, highlighting the need for further development to include broader environmental concerns, such as air pollution and climate change. Thus, future research is essential to evaluate the effectiveness of this instrument in diverse educational contexts and across wider geographic areas.

CONCLUSION

The application of the Rasch Model enables a more detailed instrument validation process, ensuring more accurate measurement compliance. Based on the data analysis, the developed instrument is proven to be unidimensional, with a Raw Variance Explained by a Measure value of 43.2%. This indicates that the instrument effectively measures the expected aspects of scientific literacy. Out of the 16 analyzed items, 13 were found to be fit and well-understood by respondents. The reliability test results show a Cronbach's Alpha value of 0.89, indicating high consistency and suggesting that the instrument can be used repeatedly. The variation in the proportion of test items in this study also demonstrates that the instrument can comprehensively assess students' abilities. The Differential Item Function (DIF) analysis results indicate that only one item shows bias, confirming that the majority of the items in this instrument are fair and do not exhibit gender bias. For future research, it is recommended that this instrument be tested on a larger population and in various educational contexts to enhance the accuracy of its generalizability. Additionally, a study is needed to evaluate the instrument's effectiveness in improving students' scientific literacy by comparing assessment results before and after its implementation in the learning process.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Public High School 1 Palangka Raya for granting permission and providing support for the implementation of this research. High appreciation is also extended to the validators for their valuable input in the development of the research instrument. Furthermore, the authors sincerely appreciate the guidance, direction, and support provided by the supervising lecturer throughout the research and manuscript preparation process.

REFERENCES

Adhari, D., Santiani, Yuliani, H., & Nasir, M. (2024). Alat Ukur Literasi Sains Pada Pembelajaran IPA Integrasi Lahan Gambut. *Sistematika Literatur Review. Kappa Journal*, *8*(2), 278–285. https://doi.org/10.29408/kpj.v8i2.26075

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 131–142.

Arikunto, S. (2010). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Rineka Cipta.

Arikunto, S. (2019). *Prosedur Penelitian: Suatu Pendekatan praktik*. Rineka Cipta.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Science*. Springer.

Bybee, R. W. (2013). *The Case for STEM Education: Challenges and Opportunities*. NSTA Press.

Danielis, M., Dato, M., Zanotti, R., Biasin, E., & Veronese, M. (2025). Development and validation of the infant nursing assessment scale: Results from exploratory factor analysis and Rasch modeling. *Journal of Pediatric Nursing*, *81*, 55–62. https://doi.org/10.1016/j.pedn.2025.01.018

De Hemptinne, J. C., Kontogeorgis, G. M., Dohrn, R., Economou, I. G., Ten Kate, A., Kuitunen, S., Fele Žilnik, L., De Angelis, M. G., & Vesovic, V. (2022). A View on the Future of Applied Thermodynamics. *Industrial and Engineering Chemistry Research*, *61*(39), 14664–14680. https://doi.org/10.1021/acs.iecr.2c01906

Ding, C. (2022). Examining the context of better science literacy outcomes among U.S. schools using visual analytics: A machine learning approach. *International Journal of Educational Research Open*, *3*. https://doi.org/10.1016/j.ijedro.2022.100191

Fagerlund, J., Palsa, L., & Mertala, P. (2025). Exploration of domains of educational purpose in K-12 data literacy education research. In *Educational Research Review* (Vol. 46). Elsevier Ltd. https://doi.org/10.1016/j.edurev.2024.100663

Fuadi, H., Robbia, A. Z., Jamaluddin, J., & Jufri, A. W. (2020). Analisis Faktor Penyebab Rendahnya Kemampuan Literasi Sains Peserta Didik. *Jurnal Ilmiah Profesi Pendidikan*, *5*(2), 108–116. https://doi.org/10.29303/jipp.v5i2.122

Holzberger, D., Reinhold, S., Lüdtke, O., & Seidel, T. (2020). A meta-analysis on the relationship between school characteristics and student outcomes in science and maths–evidence from large-scale studies. In *Studies in Science Education* (Vol. 56, Issue 1, pp. 1–34). Routledge. https://doi.org/10.1080/03057267.2020.1735758

Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. In *Teaching and Teacher Education* (Vol. 151). Elsevier Ltd. https://doi.org/10.1016/j.tate.2024.104739

Kemendikbud. (2019). *Hasil PISA Indonesia 2018: Akses makin meluas, saatnya tingkatkan kualitas.* Https://Www.Kemdikbud.Go.Id/Main/Blog/2019/12/Hasil-Pisa-Indonesia-2018-Akses-Makin-Meluas-Saatnya-Tingkatkan-Kualitas.

Krisanda, Y. I., & Harjito. (2021). Penggunaan Rasch Model Untuk Analisis Instrumen Tes Literasi Sains Materi Hidrolisis Garam. *CiE*, *10*(2). http://journal.unnes.ac.id/sju/index.php/chemined

Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing.

Nisrina, N., Jufri, A. W., & Gunawan, G. (2020). Pengembangan LKPD Berbasis Blended Learning untuk Meningkatkan Literasi Sains Peserta Didik. *Jurnal Pijar Mipa*, *15*(3), 192–199. https://doi.org/10.29303/jpm.v15i3.1880

Nurlatifah, S., Supriatna, M., Yudha, E. S., & Julius, A. (2023). Pengembangan Instrumen Kecakapan Kerja Mahasiswa (Rasch Model Analisis). *GUIDENA: Jurnal Ilmu Pendidikan, Psikologi, Bimbingan Dan Konseling*, *13*(4), 792. https://doi.org/10.24127/gdn.v13i4.8604

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. OECD. https://doi.org/10.1787/b25efab8-en

OECD. (2023). *PISA 2022 Results (Volume I)*. OECD. https://doi.org/10.1787/53f23881-en

OECD. (2024). *Science performance (PISA)*. Https://Www.Oecd.Org/En/Data/Indicators/Science-Performance-Pisa.Html.

Rahmati, U., Wahdi Ginting, F., Novita, N., Sakdiah, H., Cot Tengku Nie, J., Muara Batu, K., & Aceh Utara, K. (2024). Pengembangan Instrumen Tes Fisika Berbasis Literasi Sains Pada Materi Gerak Melingkar. In *GRAVITASI Jurnal Pendidikan Fisika dan Sains* (Vol. 7, Issue 01).

Ramli, M., Susanti, B. H., & Yohana, M. P. (2022). Indonesian Students' Scientific Literacy in Islamic Junior High School. *International Journal of STEM Education for Sustainability*, *2*(1), 53–65. https://doi.org/10.53889/ijses.v2i1.33

Rosana, D., Widodo, E., Setianingsih, W., Setyawarno, D., Hestiana, & Setiadi, Y. (2020). Pengembangan Instrumen Penilaian Model PISA Untuk Mengukur Kemampuan Literasi Sains Peserta Didik SMP. *JPSI*.

Roshid, M. M., & Haider, M. Z. (2024). Teaching 21st-century skills in rural secondary schools: From theory to practice. *Heliyon*, *10*(9). https://doi.org/10.1016/j.heliyon.2024.e30769

Saraiva, E., Silva, S., Castro, J., Araújo, D., Almeida, C., & Azevedo, M. M. (2025). Students' summer internships in a research centre: The impact on scientific literacy and the choice of a career in the STEM fields. *Heliyon*, *11*(4). https://doi.org/10.1016/j.heliyon.2025.e42472

Setiawan, A. R. (2019). Menyusun instrumen penilaian untuk pembelajaran topik lingkungan berorientasi literasi saintifik. *Prosiding Seminar Nasional Fisika*, 7–14.

Sheptian, R., Ocy, D. R., Triana, D. D., Dwi, L., & Wahyuni, U. (2024). *JISAE (Journal of Indonesian Student Assessment and Evaluation) Developing a Validated Essay-Based Assessment Instrument to Measure Science Literacy in Energy Topics*. *10*(2). https://doi.org/10.21009/JISAE

Soeharto, S., Martono, M., Hairida, H., Akhmetova, A., Arifiyanti, F., Benő, C., & Charalambos, C. (2024). The metacognitive awareness of reading strategy among pre-service primary teachers and the possibility of rating improvement using Rasch analysis. *Studies in Educational Evaluation*, *80*. https://doi.org/10.1016/j.stueduc.2023.101319

Sugiyono. (2012). *Metode Penelitian Pendekatan Kuantitatif, Dan R&D*. Alfabeta.

Sugiyono. (2014). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch: untuk Penelitian Ilmu-Ilmu Sosial* (B. Trim, Ed.). Trim Komunikata.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Trim Komunikata.

Supriadi, Sani, A., & Putra Setiawan, I. (2020). Integrasi Nilai Karakter dalam Pembelajaran Keterampilan Menulis Siswa. *Journal of Management*, *3*(3), 84–93. https://doi.org/10.2568/yum.v3i3.778

Sutrisna, N. (2021). *Analisis Kemampuan Literasi Sains Peserta Didik Sma Di Kota Sungai Penuh*. *1*(12), 2683.

Thiagarajan, S., & Semmel, S. (1974). *Cercular Model of R&D*. Leadership Training.

Tudor, G. (2023). The Interdisciplinary Approach of Physics and Its Role in Environmental Protection. *EGU General Assembly 2023*, 24–28.

Wahyuni, A. (2022). Detection of Gender Biased Using Dif (Differential Item Functioning) Analysis on Item Test of School Examination Yogyakarta. In *Jurnal Evaluasi Pendidikan* (Vol. 13, Issue 1).

Wilsa, A. W., Rusilowati, A., Susilaningsih, E., Jaja, J., & Nurpadillah, V. (2023). Validity, reliability, and item characteristics of cell material science literacy assessment instruments. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(2), 177–188. https://doi.org/10.21831/pep.v27i2.61577

Yusmar, F., & Fadilah, R. E. (2023). Analisis Rendahnya Literasi Sains Peserta Didik Indonesia: Hasil PISA Dan Faktor Penyebab. *LENSA (Lentera Sains): Jurnal Pendidikan IPA*, *13*(1), 11–19. https://doi.org/10.24929/lensa.v13i1.283