KPJ 10(2) (2026)

KAPPA JOURNAL
Physics & Physics Education

https://e-journal.hamzanwadi.ac.id/index.php/kpj/index

Development and Validation of a Science Literacy-Based Critical Thinking Assessment Instrument for Renewable Energy Content

Alifya Crystal Jasmine Zafira Zahra Budimansyah[1], Mukhayyarotin Niswati Rodliyatul Jauhariyah[2*]
[1,2]Physics Education Study Program, Universitas Negeri Surabaya, Surabaya, Indonesia.

Received: 07 June 2026
Revised: 10 July 2026
Accepted: 28 August 2026

Corresponding Author:
Mukhayyarotin Niswati Rodliyatul Jauhariyah
mukhayyarotinjauhariyah@unesa.ac.id

DOI:
https://doi.org/10.29408/kpj.v10i2.35297

Abstract: This study aims to describe the theoretical and empirical validity of a science literacy-based critical thinking test instrument on the topic of renewable energy. The developed instrument integrates science literacy competencies to measure critical thinking skills while simultaneously fostering students' science literacy within a single instrument. The instrument development followed the ADDIE model, focusing on three main stages: Analysis, Design, and Development. The instrument consists of 20 open-ended items based on Facione's critical thinking indicators, integrated with science literacy competencies. Theoretical validity was obtained through a validation sheet reviewed by three expert validators in their respective fields, while empirical validity was obtained through a pilot test involving 34 students. The results of theoretical validity for the content, construct, and language aspects were 97%, 94%, and 90%, respectively, with an overall average of 94%, categorized as highly valid. Regarding empirical validity, all test items were deemed valid, with 14 items (70%) categorized as very high and 6 items (30%) categorized as high, and they exhibited a very high level of reliability. Based on difficulty level, 11 items (55%) were categorized as moderate and 9 items (45%) as easy. The discriminant validity analysis showed that 17 items (85%) were classified as adequate and 3 items (15%) as good. Overall, all items deemed valid can be used to measure students' critical thinking skills in a comprehensive and structured manner.

Keywords: Critical Thinking; Science Literacy; Test Instruments; Renewable Energy; Validity

Introduction

The 21st century is marked by rapid advancements in science and technology. These rapid advancements have brought about significant changes in various aspects of human life (Mahrunnisya 2023). 21st-century learning is student-centered, aiming to equip learners with thinking skills such as: (1) critical thinking, (2) problem-solving, (3) metacognition, (4) communication, (5) collaboration, (6) innovation and creativity, and (7) information literacy (Mardiyah et al., 2021). In the era of the Fourth Industrial Revolution, characterized by transformation across all scientific fields, 21st-century learning is driven by several skills, such as communication, collaboration, critical thinking, problem-solving, and creativity. These skills are crucial for addressing the challenges posed by the Fourth Industrial Revolution (Lase 2019). Among them, critical thinking is one of the most crucial skills in the field of education, critical thinking is an essential skill that students must master; students who possess this skill will be better able to grasp concepts and tackle real-life problems (Ariadila et al. 2023).

Through the Merdeka Curriculum, the Indonesian education system identifies critical thinking

How to Cite:
Budimansyah, A. C. J. Z. Z. & Jauhariyah, M. N. R. (2026). Development and Validation of a Science Literacy-Based Critical Thinking Assessment Instrument for Renewable Energy Content. *Kappa Journal, 10*(2), 109-117. https://doi.org/10.29408/kpj.v10i2.35297

as one of the essential elements of the Pancasila Student Profile that must be developed (Rosmana et al. 2022). This ability does not emerge instantly but must be developed through practice and habituation by engaging students in real-world situations that require critical problem-solving (Hadiryanto and Thaib 2016). Critical thinking is the ability to evaluate information systematically and rationally to make evidence-based decisions (Mulyani 2022). Critical thinking skills can be reflected through students' ability to interpret, analyze, infer, evaluate, and explain information in solving problems (Ramadhani and Nurita 2022). Critical thinking refers to students' ability to interpret information, analyze arguments, draw conclusions, evaluate evidence, and provide logical explanations in solving problems (Calma and Davies 2026).

Critical thinking is a cognitive skill that involves actively engaging in cognitive processes to systematically analyze, evaluate, and make decisions based on available information and evidence (Juliyantika and Batubara 2022). Students with critical thinking skills tend to be better able to solve difficult problems and make sound decisions in various learning situations. Recent research also shows that STEM problem-based contextual learning supported by an e-module can improve students' critical thinking skills in science learning (Pertiwi et al. 2019). In addition, problem-based learning combined with Socratic dialogue has been shown to effectively enhance students' critical thinking skills (Pitorini 2024). Therefore, critical thinking education must be integrated into the educational curriculum to prepare a flexible and intelligent generation (Paul and Elder 2019), and to enable students to become independent individuals capable of evaluating the credibility of information in an era of information abundance (Zhao et al. 2016).

Based on various international and national studies, Indonesian students' critical thinking skills remain low (Fuad 2017). The 2018 PISA results ranked Indonesia 74th out of 79 participating countries, with scores in reading literacy, mathematics, and science still below the OECD average. The critical thinking skills of high school students in physics learning are in the very low category with an average score of 27.80%, particularly in the indicators of inference (24.00%) and evaluation (28.00%); therefore, more innovative and interactive learning strategies are needed to improve critical thinking skills sustainably (Ahzari and Akmam 2025).

The low level of critical thinking in physics education is caused by several interrelated factors. Physics education in Indonesia tends to be monotonous and dominated by teacher-led lectures, while laboratory activities are rarely conducted (Cahyana et al. 2017). Learning that is dominated by memorization may limit students' opportunities to develop higher-order thinking skills because such skills require conceptual understanding, analytical reasoning, and problem-solving activities. One of the main obstacles is the limited implementation of learning activities that engage students in scientific inquiry and problem solving, even though such activities are needed to train students' analysis, evaluation, and decision-making skills (Ahzari and Akmam 2025).

Students' science literacy competencies are closely linked to their critical thinking skills (Cahyana et al. 2017). According to the OECD (2023), science literacy is a competency developed through science education to engage with scientific issues as reflective and critical citizens in decision-making, encompassing three main competencies: explaining phenomena scientifically, designing and evaluating scientific investigations, and researching and using scientific information for evidence-based decision-making. Students with high levels of science literacy tend to have better critical thinking skills, and there is a significant positive correlation between the two (Rahayuni 2016). This relationship is supported by Listiani and Susilo (2022) who found that scientific literacy is related to critical thinking among prospective teacher students. Kusuma (2023) also reported a positive relationship between biological literacy and critical thinking skills among senior high school science students. Similarly, Apriliana and Anggrella (2024) found that students' science literacy has a positive relationship with their critical thinking skills, indicating that strengthening science literacy can contribute to the improvement of critical thinking.

Subjective essay-based tests are considered more effective for measuring critical thinking skills because they allow students to demonstrate their analytical, inferential, and evaluative abilities in a systematic and organized manner (Calma and Davies 2026). Open-ended questions are considered effective for assessing students' critical thinking skills because they allow students to construct explanations, provide reasons, and demonstrate their thinking processes. The development of test instruments is commonly conducted through Research and Development (R&D) using models such as ADDIE, followed by expert validation and empirical testing to determine the quality of the instrument (Sarwanto, Fajari, and Chumdari 2021).

Based on observations and needs analysis conducted through interviews with physics teachers at IPIEMS High School in Surabaya, it was found that the availability of assessment instruments specifically designed to measure students' critical thinking skills remains very limited. The instruments used to date have focused primarily on assessing basic knowledge, with questions sourced from the internet or textbooks. The renewable energy material included in the Merdeka Curriculum for 10th-grade physics is highly relevant to daily life and global energy issues that have been the subject of intense attention over the past few decades. Through learning this material, students are expected to be able to contribute innovative ideas and concepts to solve energy problems in the 21st century. Given the limited availability of science literacy-based assessment instruments that are theoretically and empirically validated to measure critical thinking skills in renewable energy topics, this study aims to develop a science literacy-based critical thinking assessment instrument to measure high school students' critical thinking skills in renewable energy content.

Method

This study employs the ADDIE development model as described by Branch in Sugiyono (2017). The research development process consists of five stages, as shown in Figure 1.

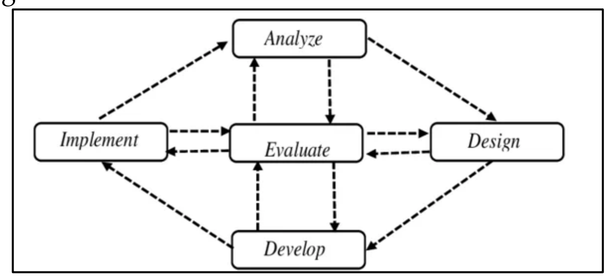

Figure 1. Diagram of the ADDIE research and development steps (Sugiyono, 2017)

The present study was limited to the Analysis, Design, and Development phases of the ADDIE model because its focus was on describing the theoretical and empirical validity of the developed test instrument. Data collection was conducted using two methods, a questionnaire and a test. The questionnaire method was implemented through a validation process conducted by experts, consisting of two lecturers from the Department of Physics at Surabaya State University and one high school physics teacher from IPIEMS Surabaya. Meanwhile, the test method was carried out through a limited trial involving 34 eleventh-grade students (Phase F) at IPIEMS High School in Surabaya. The collected data was subsequently analyzed using quantitative descriptive analysis techniques.

The analysis phase was carried out through a needs analysis to identify the level of need for the instrument to be developed. The design phase involved developing the test instrument package, which included a test blueprint, test items, answer keys, and scoring guidelines. Next, in the development stage, validity testing was conducted by experts to obtain data on the theoretical validity of the instrument. In addition, during the same stage, a limited pilot test was also conducted with students to obtain empirical validity data for each item that had been developed.

Theoretical validity refers to a situation in which an instrument is deemed acceptable by experts because it meets validity criteria based on logical reasoning—that is, it is designed in accordance with established theories and guidelines (Arikunto 2013). Generally, theoretical validity is assessed based on three main aspects: content, construct, and language. The validity percentage for each indicator is calculated using Equation 1.

$$\text{Validity Percentage} = \frac{\text{score obtained}}{\text{maximum score}} \times 100\% \qquad (1)$$

The validation sheets used generally take the form of a Likert scale with a specific score range, such as a 1–4 or 1–5 scale, in which each aspect is rated based on predetermined criteria (Riduwan 2013). After the test instrument was deemed theoretically valid, a pilot test was conducted with 34 eleventh-grade students at IPIEMS High School in Surabaya. The data from this pilot test were then analyzed to determine empirical validity, which includes item validity, reliability, difficulty level, and item discrimination. Item validity analysis was performed using the product-moment correlation equation.

$$r_{XY} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}} \qquad (2)$$

Description:
r_{XY}= Product moment correlation coefficient
N = Number of trial respondents
X = Total item score
Y = Total score

The science literacy-based test instrument developed can be considered valid if it meets the criteria with a r_{XY} value of ≥ 0.41.

Next, the instrument's reliability value is calculated using Equation 3.

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\Sigma \sigma i^2}{\sigma t^2}\right) \qquad (3)$$

Description:
r_{11} = Reliability Value
n = Number of question items
$\Sigma\sigma i^2$ = Number of variances of question item scores
σt^2 = Total variant

In assessing the reliability of a test instrument, the calculated r_{11} value is compared with the r value in the product-moment table. The instrument is considered reliable if r_{calc} is greater than r_{table}.

According to Arifin and Zainal (2016) the level of difficulty can be determined using Equation 4.

$$\text{Item Difficulty Level} = \frac{\frac{\sum x}{N}}{maximum\ score} \quad (4)$$

Description:
$\sum x$ = The number of students in each question
N = Number of students

According to Arikunto (2018) the quality of a test item is determined by its appropriate level of difficulty that is, it should not be too easy so that it remains challenging, nor too difficult so as not to dampen students' motivation.

The discriminating power of a question is the ability of a test item to distinguish students based on their ability levels (Arikunto 2018). To determine this discriminating power, according to Arifin and Zainal (2016) Equation 5 can be used.

$$DP = \frac{\sum K_A - \sum K_B}{maximum\ score} \quad (5)$$

Description:
DP = differentiating power
ΣK_A = Average of the upper group
ΣK_B = Lower Group Average

According to Arikunto (2018) a test item is considered to have discriminative power if its D value reaches or exceeds 0.20.

Result and Discussion

Analysis Phase

The analysis stage in the development of the test instrument began with a literature review and interviews. The literature review was conducted to assess the urgency of critical thinking skills and the level of their attainment among Indonesian students, while interviews were conducted with physics teachers at IPIEMS High School to obtain information regarding the use of critical thinking skills test instruments in schools. The results of this stage were used as the basis for designing the instrument to be developed.

The results of the literature review indicate that critical thinking skills are an essential competency for students, particularly in meeting the demands of 21st-century learning. These skills play a role in helping students understand and evaluate information, as well as solve problems effectively and innovatively. However, Indonesian students' critical thinking skills remain relatively low, as evidenced by various international and national studies indicating that Indonesian students' critical thinking abilities are still relatively low (Fuad 2017). This is further supported by the 2018 PISA results, which ranked Indonesia 74th out of 79 countries, with scores in reading literacy, mathematics, and science remaining below the OECD average (OECD 2019).

Based on interviews with physics teachers at IPIEMS High School, it was found that specific critical thinking skills assessment instruments are not yet available, and their use remains very limited. Although the questions used are designed to assess reasoning skills, the measurement of critical thinking skills has not been conducted specifically. This indicates that the development of critical thinking skills assessment instruments is urgently needed to support a more comprehensive learning evaluation process.

Design Phase

In this phase, a science literacy-based critical thinking skills test instrument on the topic of renewable energy was designed. The first step in the design process was to determine the question formats to be used. The selection of question formats was based on the skills to be measured, namely critical thinking skills. Calma and Karunarathne (2025) state that assessing critical thinking is more appropriately done using open-ended writing tasks because they can effectively identify various components of critical thinking. Additionally, Lestari and Widodo (2021) note that essay questions provide students with the opportunity to formulate answers based on their own thinking and to practice systematic thinking skills. Therefore, in this study, the test instrument was designed in the form of essay questions. The developed instrument features contextual stimuli aligned with science literacy competencies, namely explaining phenomena scientifically, designing and evaluating scientific investigations, and interpreting scientific data and evidence. The stimuli are presented in the form of short passages accompanied by images, graphs, or diagrams as sources of information, while also containing scientific knowledge to broaden students' understanding of renewable energy (Masithoh and Jauhariyah 2024).

Based on these stimuli, test items were then developed that align with the indicators of critical thinking skills. The development of test items was guided by the five indicators of critical thinking identified by Facione: interpretation, analysis, evaluation, explanation, and inference (Haryanti, Susongko, and Arfiani 2024). Essay-based test instruments are considered appropriate for assessing critical thinking because they allow students to present reasoning, explanations, and argumentation in a more open and structured manner (Ikhsan et al. 2025). Self-regulation indicators were not measured in this study because they are more appropriately measured using non-test instruments such as questionnaires or observations (Yusmar and Fadilah 2023). Each test item was designed to measure one critical thinking indicator integrated with science literacy competencies.

In this study, each subtopic contains three indicators of science literacy competencies, each of which was developed into one test item. Thus, each subtopic yielded three test items. The renewable energy material used covers seven subtopics, namely the urgency of electricity needs, forms of energy, energy conversion and the law of conservation of energy, energy sources, renewable and non-renewable energy, the impact of energy exploration, and the utilization of local energy potential. In total, 20 test items were developed.

The test instruments developed include answer keys and scoring guidelines. The specifications of the instruments developed are presented in Table 1.

Table 1. Specification of the Developed Test Instrument

Subtopic: Energy Conversion and the Law of Conservation of Energy
Science Literacy Competency: Explaining phenomena scientifically
Curriculum Standards Indicator: Explanation

Question

An engineering laboratory conducted tests on the braking systems of two motorcycles with the same mass. The two motorcycles were ridden at different speeds, then braked until it came to a complete stop. The test results are recorded as follows:

Motorcycle	Initial Speed	Initial Temperature	Final Temperature
Motorcycle P	12 km/h	27°C	34°C
Motorcycle Q	36 km/h	27°C	55°C

Source: Karunaratne, A. and Widanowo, Y. (2009). *Energy and Its Applications.*

Based on the data above, explain why the brakes on Motor Q experience a higher temperature increase than those on Motor P after braking! Relate your explanation to the concept of kinetic energy and the law of conservation of energy!

Answer Key	Score
The brakes on Motor Q get hotter because it has a higher initial speed (36 km/h), so its kinetic energy is greater than that of Motor P (12 km/h). Based on the formula Ek = ½mv², Motor Q, which has three times the speed of Motor P, has 9 times more kinetic energy (since speed is squared). When braked, the motor's kinetic energy is converted into heat energy due to friction between the brake pads and the brake disc. The greater the kinetic energy, the more heat energy is produced, so the temperature of Motor Q's brakes increases by 28°C (from 27°C to 55°C), while Motor P's temperature increases by only 7°C (from 27°C to 34°C). This is consistent with the law of conservation of energy, which states that energy is not lost but is converted from kinetic energy into heat and sound energy.	5
Mentioning 4 out of 5 aspects clearly and correctly	4
Mentioning 3 out of 5 aspects, typically: higher speed → greater kinetic energy, which is converted into heat, and there are numerical data, but no mention of the formula or the law of conservation of energy	3
It mentions only two aspects: Motor Q is faster and generates more heat, without explaining the concept of kinetic energy, the formula, or the law of conservation of energy	2
Mentioning only one aspect or providing an incomplete or irrelevant answer	1

Development Phase

The developed test instrument was validated by three experts, including two university lecturers and one teacher. The validation covered content, construct, and language aspects, with the results presented in Table 2.

Table 2. Theoretical Validity Results

No	**Assessment Aspects**	**Validity Value (%)**	**Categories**
1.	Material	97%	Highly Valid
2.	Construction	94%	Highly Valid
3.	Language	90%	Highly Valid
Total		**281%**	
Average		**94%**	**Highly Valid**

The results of the expert evaluation indicate that the science literacy-based critical thinking test instrument on renewable energy is in the "highly valid" category, with an average validity percentage of 94%. This indicates that the developed instrument is suitable for use in research. From the content perspective, the instrument achieved a validity score of 97%, falling into the "highly valid" category. This indicates that the test items were designed in alignment with learning outcomes, learning objectives, and indicators that represent the overall content. The alignment between the instrument's content and the learning material is one of the primary criteria for determining an instrument's validity (Arikunto 2018).

In terms of construction, a percentage of 94% was obtained, falling into the "highly valid" category. This indicates that the item construction adheres to sound principles, including clear instructions, complete supporting components, and alignment with scoring guidelines (Sugiyono 2022). Meanwhile, the language aspect received a validity score of 90%, falling into the "highly valid" category. This indicates that the language used in the instrument adheres to the principles of good and communicative Indonesian, thereby avoiding ambiguity.

The instrument, which had been deemed theoretically sound, was then pilot-tested on a limited scale with 34 high school students at IPIEMS. The students' test results were subsequently analyzed to determine the instrument's empirical validity. The instrument was deemed to meet empirical criteria if it covered the aspects of item validity, reliability, difficulty level, and discriminative power with a minimum rating of "adequate" (Arikunto 2018).

The high content validity score may be attributed to the systematic alignment between science literacy competencies, critical thinking indicators, and renewable energy content. Similar findings were reported by Haryanti et al. (2024), who found that critical thinking instruments developed using Facione's framework achieved high expert validity due to clear indicator-item alignment.

Thus, the developed instrument is not only theoretically sound but also has the potential to provide accurate measurement results in identifying students' critical thinking skills. The results of the analysis are presented in Table 3.

Table 3. Valid Test Items Based on Theoretical and Empirical Validity

Question	VT	VE			Description
		VEBS	DI	DP	
1	94%	0.830	0.888	0.333	Valid
2	94%	0.827	0.858	0.333	Valid
3	94%	0.827	0.870	0.311	Valid
4	94%	0.793	0.882	0.289	Valid
5	94%	0.842	0.858	0.311	Valid
6	94%	0.853	0.852	0.356	Valid
7	94%	0.838	0.858	0.311	Valid
8	94%	0.840	0.706	0.356	Valid
9	94%	0.851	0.700	0.400	Valid
10	94%	0.855	0.700	0.422	Valid
11	94%	0.862	0.730	0.378	Valid
12	94%	0.843	0.700	0.378	Valid
13	94%	0.873	0.682	0.400	Valid
14	94%	0.849	0.682	0.400	Valid
15	94%	0.642	0.448	0.288	Valid
16	94%	0.736	0.448	0.400	Valid
17	94%	0.663	0.424	0.311	Valid
18	94%	0.833	0.424	0.444	Valid
19	94%	0.712	0.436	0.444	Valid
20	94%	0.686	0.394	0.356	Valid

VT=Theoretical validity; VE=Empirical Validity; VEBS=Empirical Validity Based on Student Responses; DI=Difficulty Index; DP=Discrimination Power; VE consists of VEBS, DI, and DP.

Based on Table 3, it is evident that all 20 test items were deemed valid based on the results of empirical validity. The overall reliability coefficient of the instrument was 0.970 (Cronbach's Alpha), indicating very high reliability. This aligns with the results of theoretical validity, which also indicated high validity scores for all test items. This means that the test items, which had been deemed valid by experts, were proven to remain consistently valid when administered directly to students. Nevertheless, it is important to understand that alignment between theoretical and empirical validity does not always occur in every instrument development. Low empirical validity scores may be related to item characteristics, such as inappropriate difficulty levels, weak discrimination power, and poorly functioning distractors, which can reduce the quality of the test instrument (Ramadhan et al. 2020).

The test items that have been declared valid are then grouped based on critical thinking ability indicators. The results of the grouping of test items by critical thinking ability indicators are presented in Table 4.

Table 4. Valid Test Items for Each Critical Thinking Indicator

Critical Thinking Indicators	Question Items Before the Trial	Question Items After the Trial	
		Valid	Invalid
Interpretation	2, 14	2, 14	-
Analysis	1, 8, 18, 19, 20	1, 8, 18, 19, 20	-
Evaluation	3, 5, 6, 11, 12, 15, 17	3, 5, 6, 11, 12, 15, 17	-
Explanation	4, 7, 10, 13, 16	4, 7, 10, 13, 16	-
Inference	9	9	-

The empirical validity of the test items was determined by analyzing student responses using the product-moment correlation equation. Based on the analysis of the 20 developed test items, all items were deemed valid, with 14 items (70%) falling into the very high category ($r \geq 0.81$) and 6 items (30%) falling into the high category ($0.61 \leq r \leq 0.80$). No items fell into the adequate, low, or very low categories. Test items deemed valid based on empirical item validity are those whose scores are comparable to the overall test score (Arikunto 2018). Thus, all 20 developed test items were empirically validated and capable of measuring the targeted competencies. This aligns with the statement Arikunto (2018) that an item is considered valid if it can measure the skill intended to be measured. The high empirical validity scores for the majority of items indicate that each item has a strong correlation with the total test score, meaning the developed instrument possesses good empirical quality and is suitable for measuring students' abilities.

Reliability refers to the degree of consistency of an instrument. Reliability can be determined after the test instrument has been pilot-tested (Arikunto 2018). Based on the analysis of student responses using the Cronbach's Alpha formula, a reliability value of 0.970 was obtained; with this value, the instrument is categorized as reliable, meeting the criteria for very high reliability. This is evidenced by the calculated r value (0.970) being greater than the table of r value (0.339) for a sample size of 34 students, thereby establishing the instrument's statistical reliability. These results indicate that the developed test instrument exhibits high consistency, meaning that when administered to students at different times, it will yield relatively similar results. In line with this, Lina, Helendra and Asih (2018) argue that a test instrument is categorized as reliable if the test results show stability. This means that students will be ranked in the same order even if the instrument is administered at different times.

The difficulty level indicates how difficult or easy a test item is for test-takers. The difficulty level can be determined based on the proportion of students who are able to answer each item correctly (Arikunto 2018). The difficulty levels of the test items developed in this study vary. Based on the analysis results, out of the 20 test items administered, 11 items (55%) fell into the moderate category and 9 items (45%) fell into the easy category. No test items were found to fall into the difficult category; thus, the difficulty level distribution of the instrument was dominated by the moderate category, followed by the easy category. A test item is categorized as easy if almost all students are able to answer it correctly, while a test item is categorized as difficult if only a few or even no students can answer it correctly (Rajagukguk 2023).

The dominance of moderately difficult items indicates that the instrument is appropriately challenging for students and capable of eliciting critical thinking processes without causing excessive cognitive burden. According to (Arikunto 2018), questions that are too easy will not encourage the improvement of students' thinking and problem-solving skills, while questions that are too difficult have the potential to cause students to feel pessimistic about answering the questions. Therefore, a balanced composition of difficulty levels between the easy, moderate, and difficult categories is essential in the development of a high-quality test instrument.

Item discrimination refers to the ability of a test item to distinguish between high-ability and low-ability students (Arikunto and Suharsimi 2021). The discrimination index is an important component of item analysis because it indicates how well an item differentiates students with different levels of achievement (Kim et al. 2023). A positive discrimination index shows that students in the upper group answer the item correctly more often than those in the lower group, indicating that the item functions properly (Rezigalla et al. 2024). This aligns with the view Arikunto and Suharsimi (2021) that if the number of subjects is less than 100 students, the division into top and bottom groups is conducted with each group comprising 50% of the total test participants. The absence of poor discrimination items suggests that all questions functioned effectively in differentiating students with varying levels of critical thinking ability.

The very high reliability coefficient indicates strong internal consistency among items. This finding is consistent with previous studies reporting that essay-based instruments with well-defined scoring rubrics can produce highly reliable measurements of higher-order thinking skills. This finding is consistent with Ikhsan et al. (2025), who reported that essay-based critical thinking instruments supported by detailed scoring rubrics tend to demonstrate high reliability.

Based on the analysis of the 20 test items, 17 items (85%) fell into the "adequate" category and 3 items (15%) into the "good" category. No items were classified as "very good," "poor," or "very poor"; thus, all items demonstrated adequate discriminative power, and the instrument was deemed suitable for use. According to Arikunto & Suharsimi (2021), a question is considered to lack good discriminative power if it can be answered correctly by both groups simultaneously, or conversely, cannot be answered by either group, meaning the question fails to accurately reflect differences in students' abilities.

The adequate discrimination indices indicate that the instrument is capable of distinguishing students with different levels of critical thinking ability, which is an important characteristic of a quality assessment instrument.

This study has several limitations. The empirical validation was conducted using a limited sample of 34 students from one senior high school. In addition, the study focused only on validity, reliability, difficulty level, and discrimination power, without examining construct validity through factor analysis or conducting large-scale field testing. Future studies should involve larger and more diverse samples to further establish the psychometric quality of the instrument.

Conclusion

The results of theoretical validity for content, construct, and linguistic aspects were 97%, 94%, and 90%, respectively, with an overall average of 94%, categorized as highly valid. Regarding empirical validity, all 20 test items were deemed valid, with 14 items (70%) categorized as very high and 6 items (30%) categorized as high. Thus, the science literacy-based critical thinking test instrument developed was deemed suitable, and all 20 test items can be used to measure students' critical thinking skill profiles. Therefore, the developed instrument can be used as an alternative assessment tool for measuring students' critical thinking skills while simultaneously promoting science literacy in renewable energy learning and supporting classroom assessment practices in senior high schools.

Acknowledgments

The authors would like to express their sincere gratitude to all parties who supported and contributed to the completion of this study. Special appreciation is extended to the Physics Education Study Program, Universitas Negeri Surabaya, and SMA IPIEMS Surabaya for their support, guidance, and assistance throughout the research process.

References

Ahzari, S., & Akmam, A. (2025). Analyzing students' critical thinking as a basis for developing interactive physics multimedia with generative learning and cognitive conflict strategies. *Jurnal Pendidikan Fisika, 13*(2), 163–176. https://doi.org/10.26618/jpf.v13i2.17702

Ariadila, S. N., Silalahi, Y. F. N., Fadiyah, F. H., Jamaludin, U., & Setiawan, S. (2023). Analisis pentingnya keterampilan berpikir kritis terhadap pembelajaran bagi siswa. *Jurnal Ilmiah Wahana Pendidikan, 9*(20), 664–669. https://doi.org/10.5281/zenodo.8436970

Arifin, Z. (2016). *Evaluasi pembelajaran*. Bandung, Indonesia: PT Remaja Rosdakarya.

Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan*. Jakarta, Indonesia: Bumi Aksara.

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan*. Jakarta, Indonesia: Bumi Aksara.

Cahyana, U., Paristiowati, M., Savitri, D. A., & Hasyrin, S. N. (2017). Developing and application of mobile game based learning (M-GBL) for high school students' performance in chemistry. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(10), 7037–7047. https://doi.org/10.12973/ejmste/78728

Calma, A., & Davies, M. (2026). Assessing students' critical thinking abilities via a systematic evaluation of essays. *Studies in Higher Education, 51*(2), 422–437. https://doi.org/10.1080/03075079.2025.2470969

Fuad, N. M., Zubaidah, S., Mahanal, S., & Suarsini, E. (2017). Improving junior high schools' critical thinking skills based on test three different models of learning. *International Journal of Instruction, 10*(1), 101–116. https://doi.org/10.12973/iji.2017.1017a

Hadiryanto, S., & Thaib, D. (2016). Peningkatan kemampuan berpikir kritis siswa SMP melalui pembelajaran berbasis masalah pada konsep respirasi. *EduHumaniora: Jurnal Pendidikan Dasar, 8*(1), 55–65. https://doi.org/10.17509/eh.v8i1.5122

Haryanti, H., Susongko, P., & Arfiani, Y. (2024). Pengembangan instrumen penilaian berpikir kritis menurut Facione pada pembelajaran IPA di sekolah menengah pertama berbasis model Rasch. *PSEJ: Pancasakti Science Education Journal, 9*(2), 78–87. https://doi.org/10.24905/psej.v9i2.207

Ikhsan, M., Atun, S., Agusta, F., Unayah, H., Buhera, R., Pamungkas, O., Sarip, M., & Sitorus, P. A. (2025). Development of critical thinking essay test instrument and prosocial intention questionnaire for environmental care in students. Journal Evaluation in Education, 6(1), 66–78. https://doi.org/10.37251/jee.v6i1.1273

Juliyantika, T., & Batubara, H. H. (2022). Tren penelitian keterampilan berpikir kritis pada jurnal pendidikan dasar di Indonesia. *Jurnal Basicedu, 6*(3), 4731–4744. https://doi.org/10.31004/basicedu.v6i3.2869

Kim, Y. H., Kim, B. H., Kim, J., Jung, B., & Bae, S. (2023). Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2022, Korea: A psychometric study. Journal of Educational Evaluation for Health Professions, 20, 31. https://doi.org/10.3352/jeehp.2023.20.31

Kusuma, A. S. H. M. (2023). Correlation in biological literacy and critical thinking skills of science students through problem-based learning model integrated by Macromedia Flash. Jurnal Pijar MIPA, 18(3), 369–375. https://doi.org/10.29303/jpm.v18i3.4928

Lase, D. (2019). Pendidikan di era revolusi industri 4.0. *Sundermann: Jurnal Ilmiah Teologi, Pendidikan, Sains, Humaniora dan Kebudayaan, 12*(2), 28–43. https://doi.org/10.36588/sundermann.v1i1.18

Lestari, P., & Widodo, A. (2021). Developing critical thinking skills through science literacy-based learning. *Journal of Physics: Conference Series, 1806*, 012106. https://doi.org/10.1088/1742-6596/1806/1/012106

Lina, Y. R., Helendra, H., & Arsih, F. (2018). Pengembangan instrumen asesmen berbasis literasi sains tentang materi sistem pencernaan makanan, zat aditif, dan zat adiktif untuk SMP. *Bioeducation Journal, 2*(2), 145–155. https://doi.org/10.24036/bioedu.v2i2.73

Listiani, I., Susilo, H., & Sueb, S. (2022). Relationship between scientific literacy and critical thinking of prospective teachers. AL-ISHLAH: Jurnal Pendidikan, 14(1), 721–730. https://doi.org/10.35445/alishlah.v14i1.1355

Mahrunnisya, D. (2023). Keterampilan pembelajar di abad ke-21. *JUPENJI: Jurnal Pendidikan Jompa Indonesia, 2*(1), 101–109. https://doi.org/10.57218/jupenji.Vol2.Iss1.598

Mardhiyah, R. H., Aldriani, S. N. F., Chitta, F., & Zulfikar, M. R. (2021). Pentingnya keterampilan belajar di abad 21 sebagai tuntutan dalam pengembangan sumber daya manusia. *Lectura: Jurnal Pendidikan, 12*(1), 29–40. https://doi.org/10.31849/lectura.v12i1.5813

Masithoh, N. D., & Jauhariyah, M. N. R. (2024). Analisis profil kompetensi literasi sains peserta didik yang diukur menggunakan instrumen berbasis kelas pada materi energi terbarukan. *Inovasi Pendidikan Fisika, 13*(3), 184–190. https://doi.org/10.26740/ipf.v13n3.p184-190

Mulyani, A. Y. (2022). Pengembangan critical thinking dalam peningkatan mutu pendidikan di Indonesia. *DIAJAR: Jurnal Pendidikan dan Pembelajaran, 1*(1), 100–105. https://doi.org/10.54259/diajar.v1i1.226

OECD. (2019). *PISA 2018 results (Volume I): What students know and can do*. Paris, France: OECD Publishing. https://doi.org/10.1787/5f07c754-en

OECD. (2023). *PISA 2025 science framework*. Retrieved from https://pisa-framework.oecd.org/science-2025/

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking: Concepts and tools*. Tomales, CA: Foundation for Critical Thinking.

Pertiwi, N. P., Saputro, S., Yamtinah, S., & Kamari, A. (2024). Enhancing critical thinking skills through STEM problem-based contextual learning: An integrated e-module education website with virtual experiments. Journal of Baltic Science Education, 23(4), 739–766. https://doi.org/10.33225/jbse/24.23.739

Pitorini, D. E., Suciati, S., & Harlita, H. (2024). Students' critical thinking skills using an e-module based on problem-based learning combined with Socratic dialogue. Journal of Learning for Development, 11(1), 52–65. https://doi.org/10.56059/jl4d.v11i1.1014

Rahayuni, G. (2016). Hubungan keterampilan berpikir kritis dan literasi sains pada pembelajaran IPA terpadu dengan model PBM dan STM. *Jurnal Penelitian dan Pembelajaran IPA, 2*(2), 131–146. https://doi.org/10.30870/jppi.v2i2.926

Rajagukguk, M. J. T., & Naibaho, D. (2023). Mampu memilih soal berdasarkan tingkat kesukaran. *Jurnal Pendidikan Sosial dan Humaniora, 2*(4), 12736–12747.

Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The quality of test instruments constructed by teachers in Bima Regency, Indonesia: Document analysis. *International Journal of Instruction, 13*(2), 507–518. https://doi.org/10.29333/iji.2020.13235a

Ramadhani, F. A., & Nurita, T. (2022). The critical thinking skills of junior high school student on simple machines material. International Journal of Education and Teaching Zone, 1(2), 59–67. https://doi.org/10.57092/ijetz.v1i2.29

Rezigalla, A. A., Eleragi, A. M. S., Elhassan, A. M., & Isa, A. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. BMC Medical Education, 24, 445. https://doi.org/10.1186/s12909-024-05433-y

Riduwan. (2013). *Skala pengukuran variabel-variabel penelitian*. Bandung, Indonesia: Alfabeta.

Rosmana, P. S., Iskandar, S., Fauziah, H., Azzifah, N., & Khamelia, W. (2022). Kebebasan dalam kurikulum prototype. *AS-SABIQUN, 4*(1), 115–131. https://doi.org/10.36088/assabiqun.v4i1.1683

Sarwanto, S., Fajari, L. E. W., & Chumdari, C. (2021). Open-ended questions to assess critical-thinking skills in Indonesian elementary school. *International Journal of Instruction, 14*(1), 615–630. https://doi.org/10.29333/iji.2021.14137a

Sugiyono. (2017). *Metode penelitian dan pengembangan: Research and development*. Bandung, Indonesia: Alfabeta.

Sugiyono. (2022). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung, Indonesia: Alfabeta.

Yusmar, F., & Fadilah, R. E. (2023). Analisis rendahnya literasi sains peserta didik Indonesia: Hasil PISA dan faktor penyebab. *Lensa: Jurnal Kependidikan Fisika, 13*(1), 11–19. https://doi.org/10.24929/lensa.v13i1.283

Zhao, C., Pandian, A., Kaur, M., & Singh, M. (2016). Instructional strategies for developing critical thinking in EFL classrooms. *English Language Teaching, 9*(10), 14–21. https://doi.org/10.5539/elt.v9n10p14