JISAE
Journal of Indonesian Student Assessment and Evaluation
https://doi.org/10.21009/JISAE
JISAE (Journal of Indonesian Student Assessment and Evaluation)
ISSN : P-ISSN: 2442-4919 | E-ISSN: 2597-8934
Website : http://journal.unj.ac.id/unj/index.php/jisae
Vol 10 No 2 (2024)

Developing a Validated Essay-Based Assessment Instrument to Measure Science Literacy in Energy Topics

Riky Sheptian[*],
Universitas Negeri Jakarta

Dwi Rismi Ocy,
Universitas Negeri Jakarta

Dinny Devi Triana,
Universitas Negeri Jakarta

Lussy Dwi Utami Wahyuni,
Universitas Negeri Jakarta

ABSTRACT

Science literacy is a critical skill in the 21st century, particularly for middle school students navigating the complexities of energy concepts in science education. This study aims to develop and validate an essay-based assessment instrument designed to measure students' science literacy skills in the context of energy. Employing a mixed-methods approach, the research involved expert validation, panel validation, and empirical testing with 111 students from junior high school in Sukabumi, Indonesia. The results revealed that the instrument possesses strong content validity, with Aiken's V, CVR, and I-CVI scores exceeding 0.80. Construct validity analysis using Exploratory Factor Analysis (EFA) confirmed the suitability of the factor structure, while reliability tests yielded McDonald's ω = 0.886 and Cronbach's α = 0.922, indicating excellent internal consistency. The instrument also includes a detailed analytical rubric to ensure objective and transparent evaluation. This validated tool offers educators a reliable means of assessing students' science literacy comprehensively, addressing a critical gap in science education evaluation. By fostering more effective assessment practices, this study contributes to enhancing the quality of science education in Indonesia and beyond.

Address for Correspondence:
1. ricky01abiogenesis@gmail.com*
2. dwirismiocy@gmail.com
3. dinnydevi@unj.ac.id
4. lussysf@unj.ac.id

Keywords: Science Literacy; Assessment Instrument; Essay Test; Energy Concepts; Exploratory Factor Analysis (EFA).

INTRODUCTION

Science literacy is an essential skill for students in the era of globalization and rapid technological advancement. This ability involves not only an understanding of scientific concepts but also the capacity to apply that knowledge in everyday decision-making (Santoso et al., 2023). In the context of education, science literacy serves as a key indicator of learning success, warranting serious attention from educators.

Despite its importance, research indicates that science literacy among students in Indonesia remains low. According to the 2022 PISA results, the average science literacy score for Indonesian students is 383, significantly lower than the OECD global average of 485. Only about 34.16% of Indonesian students achieve the minimum competency level (Level 2) in science literacy, which is far below the OECD average of 75.51% (OECD, 2022). These results highlight that Indonesian students' science literacy is at a very low level on a global scale. Preliminary studies in several middle schools in the Sukabumi region also show that only around 30% of students meet the minimum competency criteria in the topic of energy, which is a crucial subject in the middle school science curriculum.

While previous studies have explored various aspects of science literacy assessment, such as the development of multiple-choice tests (Rusilowati et al., 2018) and project-based assessments (Hardjo et al., 2019), there remains a notable gap in the development of comprehensive essay-based instruments specifically designed to measure science literacy in the context of energy concepts. Existing assessment tools often focus on isolated aspects of science literacy (Fathia et al., 2023) or lack the depth needed to evaluate students' critical thinking and analytical skills (Ramadani et al., 2024). This study addresses these limitations by developing and validating an essay-based assessment instrument that integrates multiple dimensions of science literacy while incorporating robust analytical rubrics for objective evaluation. The innovative approach of combining detailed analytical rubrics with essay-based assessment provides a more nuanced and comprehensive evaluation of students' science literacy skills, particularly in understanding and applying energy concepts.

METHOD

This study employs mixed-methods with qual-QUAN design that integrates both qualitative and quantitative methods to develop and validate an assessment instrument aimed at measuring middle school students' science literacy skills in the context of energy. The research process begins with an in-depth literature review to identify relevant science literacy indicators. This review includes an analysis of curriculum documents, theoretical studies, and previous research findings. Additionally, interviews with science education experts were conducted to ensure that the generated indicators align with the context of science learning at the middle school level.

Based on the established indicators, the assessment instrument was developed through the creation of a blueprint that encompasses the indicators, the aspects being measured, and the test items. The constructed essay-type questions were designed to assess students' abilities in depth. An analytical rubric was also developed to support structured and objective assessment.

The instrument validation phase involved several steps. First, expert validation was conducted using the Delphi Test method, involving three science education experts who evaluated the appropriateness of the indicators and test items. Second, a panel validation was performed using Aiken's index, which included 11 practitioners and science teachers to assess the clarity, relevance, and feasibility of the test items. Third, content validation was carried out by calculating the Content Validity Ratio (CVR) and the Item Content Validity Index (I-CVI) to ensure the relevance and representation of each test item concerning the science literacy indicators.

The validity and reliability testing of the instrument were conducted quantitatively using Classical Test Theory (CTT). The instrument was trialed with 111 students from Al Khoiriyah Al Husna Middle School, and the data obtained were analyzed using Exploratory Factor Analysis (EFA) to evaluate construct validity and the reliability of the test items. The results of this analysis were used to revise the instrument, enhancing its validity and reliability in measuring middle school students' science literacy skills in the context of energy.

RESULTS AND DISCUSSION

1. Science Literacy Indicators for Junior High School Students

Based on the results of a literature review, curriculum analysis, and interviews with science education experts, science literacy indicators for junior high school students have been identified and categorized into three main dimensions: (1) Scientific Concept Understanding, (2) Critical Thinking, and (3) Scientific Communication (verbal). Each dimension is supported by specific indicators that reflect the expected science literacy skills. Table 1 presents these indicators.

Table 1. Science Literacy Skill Indicator

Dimension	No	Indicator
Understanding Science Concepts	1	Ability to Explain
	2	Ability to Apply
Critical Thinking	3	Evaluate Information/Investigation/Argument

Dimension	No	Indicator
	4	Analyze Information/Investigation/Argument
Scientific Communication (verbal)	5	Interpret Data
	6	Draw Conclusions Based on Data

As shown in Table 1, the formulation of these indicators was carried out through an intensive and iterative process involving discussions with science education experts. These discussions focused on ensuring that the indicators are relevant to the educational context of junior high school students and aligned with the goals of the curriculum. Experts provided critical insights into the scientific literacy demands expected at this level, including the specific skills students need to develop, such as understanding scientific concepts, engaging in critical analysis, and effectively communicating scientific findings. Their input ensured that the indicators address not only foundational knowledge but also higher-level cognitive and practical skills essential for science literacy.

These indicators were designed to serve as the basis for constructing a robust assessment instrument that comprehensively measures students' abilities in three key dimensions: (1) cognitive understanding (scientific concept mastery), (2) critical thinking (analysis and evaluation of information), and (3) communication (verbal expression of scientific reasoning). By covering these dimensions, the instrument is tailored to provide a holistic evaluation of students' science literacy, going beyond rote memorization to assess meaningful engagement with scientific concepts and practices.

In addition, the formulation process carefully integrated middle to higher-order thinking skills (C3-C5), including applying, analyzing, evaluating, as outlined in Bloom's taxonomy. This integration ensures that the indicators not only reflect basic competencies but also challenge students to engage in deeper levels of thinking and problem-solving. These higher-order skills are vital for preparing students to meet the demands of 21st-century education, which emphasizes critical thinking, creativity, and effective communication. By embedding these skills within the indicators, the framework supports the cultivation of students' abilities to apply scientific knowledge in real-world contexts, analyze complex problems, and communicate their findings with clarity and precision.

The development of the assessment instrument for measuring junior high school students' science literacy skills in the context of energy is supported by a set of clearly defined item indicators and criteria ratings, as presented in Table 2.

Table 2. Item Indicators and Criteria Rating for The Scoring Rubric

No	Item Indicators	Code	Criteria Rating	Cognitive Level
1	Explaining the concepts of potential and kinetic energy in everyday life	A1	Identifying the position of maximum potential energy	C4 (Analyzing)
		A2	Identifying the position of maximum kinetic energy	
		A3	Explaining the process of energy transformation	
2	Applying the concept of energy transformation in household appliances	B1	Applying the concept of energy transformation	C3 (Applying)
		B2	Identifying the benefits of energy-efficient rice cookers	
3	Evaluating information about the use of renewable energy sources	C1	Evaluating opinions	C5 (Evaluating)
		C2	Identifying benefits	
		C3	Identifying challenges	
4	Analyzing energy efficiency in household appliances	D1	Analyzing electricity consumption	C4 (Analyzing)
		D2	Analyzing lifespan	
		D3	Analyzing costs	
5	Interpreting energy usage data in household contexts	E1	Interpreting usage patterns	C4 (Analyzing)
		E2	Identifying contributing factors	

No	Item Indicators	Code	Criteria Rating	Cognitive Level
6	Drawing conclusions from energy conversion efficiency data for energy source development	F1	Drawing conclusions	C5 (Evaluating)
		F2	Identifying reasons for selection	
		F3	Considering additional factors	

As shown in Table 2, these indicators are designed to comprehensively evaluate students' conceptual understanding, critical thinking, and scientific communication skills.

Each item indicator corresponds to specific cognitive levels, following Bloom's Taxonomy. For instance, indicators such as explaining the concepts of potential and kinetic energy (A1–A3) fall under the Analyzing level (C4), requiring students to break down and explain the relationships between components. Similarly, the application of energy transformation concepts in household appliances (B1–B2) aligns with the Applying level (C3), where students demonstrate their ability to utilize theoretical concepts in practical contexts.

Higher-order cognitive skills are also targeted, as reflected in items related to evaluating the use of renewable energy sources (C1–C3) and drawing conclusions from energy efficiency data (F1–F3). These items align with the Evaluating level (C5), challenging students to critically assess and synthesize information.

The criteria ratings outlined in the rubric further ensure objectivity and clarity in the assessment process. For example, criteria such as Identifying the position of maximum potential energy (A1) and Analyzing electricity consumption (D1) provide structured benchmarks for evaluating students' performance. This detailed alignment between indicators, codes, criteria, and cognitive levels ensures that the instrument effectively measures the intended constructs, fostering both validity and reliability in the assessment of energy-related science literacy.

2. Delphi Method

The Delphi method is a structured communication technique that gathers insights and opinions from a panel of experts through multiple rounds of questioning, ultimately leading to a consensus on specific issues. The results obtained from the Delphi process are crucial for informing the development and validation of assessment instruments, as they provide valuable expert feedback on the relevance, clarity, and appropriateness of test items (Dijkstra et al., 2012; Goldhammer et al., 2020). By incorporating expert opinions, the assessment tool can be refined to better meet educational objectives and ensure that it effectively measures the intended constructs. This iterative process not only enhances the validity of the instrument but also fosters a sense of reliability and trustworthiness in the evaluation outcomes (Colbert-Getz et al., 2017).

Table 3. Delphi Test Results

No	Item Indicator	Expert Comments	Revisions Made	Conclusion
1	Explaining the concepts of potential and kinetic energy in everyday life	The language of the question is too technical.	Changed the language to be simpler and more suitable for students	Approved
2	Applying the concept of energy transformation in household appliances	The stimulus is good, but additional images are needed	Added images of household appliances as illustrations	Approved
3	Evaluating information about the use of renewable energy sources	The instructions are clear, but the assessment rubric needs strengthening	Added more detailed evaluation criteria in the rubric	Approved
4	Analyzing the efficiency of energy use in household appliances	The data table is relevant, but terminology needs adjustment for middle school context	Replaced technical terms with simpler language.	Approved
5	Interpreting energy usage data in the context of households	The stimulus needs to be supplemented with	Added guidance for graph interpretation in the stimulus	Approved

No	Item Indicator	Expert Comments	Revisions Made	Conclusion
		guidance for reading graphs		
6	Drawing conclusions from energy conversion efficiency data for the development of energy sources	The stimulus is too complex for middle school students	Simplified the data table and focused on the main concepts	Approved

Table 3 provides a detailed summary of the expert feedback, subsequent revisions, and final approval status for each test item. These modifications were systematically carried out to enhance the clarity, accessibility, and overall validity of the assessment instrument.

For Item 1, the content was revised by simplifying technical language to align with the cognitive abilities of middle school students, ensuring the question remains comprehensible without compromising its scientific accuracy. Item 2 was enriched with visual aids, such as illustrations of household appliances, to create stronger connections between the assessment and real-life contexts, thereby increasing its relevance.

The scoring rubric for Item 3 was refined to include detailed and specific criteria for evaluating students' skills in analyzing and evaluating renewable energy sources, thus ensuring consistency and objectivity in assessment. For Item 4, technical terminology within the accompanying data table was replaced with simplified language, making the task more approachable for the target age group.

To support students' interpretation and analytical skills, Item 5 was revised by including clear instructions for interpreting graphs, enabling students to engage with the visual data more effectively. Finally, Item 6 underwent adjustments to reduce the complexity of the data table, focusing only on essential concepts to ensure students can meaningfully interact with the task without unnecessary cognitive overload.

These strategic revisions collectively enhance the instrument's construct validity, content validity, and alignment with students' cognitive development. By refining the design based on expert input, the instrument assessment is optimized to reliably measure scientific literacy skills while maintaining high usability and practicality in educational settings.

3. Content Validity Test

Content validity is a critical aspect of assessment development, as it ensures that the test items accurately represent the construct being measured and are relevant to the specific educational objectives (Anggara & Abdillah, 2023; Nikmard & Mohamadi Zenouzagh, 2020). Establishing content validity involves gathering expert opinions to evaluate whether the assessment items adequately cover the intended content area and align with the learning goals. This process not only enhances the credibility of the assessment tool but also ensures that it effectively measures what it is designed to assess (Gasmalla & Tahir, 2021; Giraldo et al., 2023). By confirming that the items are relevant and necessary, content validity contributes to the overall reliability and effectiveness of the evaluation, ultimately supporting better educational outcomes.

Table 4. Content Validity Results

No	Item Indicator	Aiken V	CVR	I-CVI
1	Explaining the concepts of potential and kinetic energy	0.954	1.00	1.00
2	Applying the concept of energy transformation in household appliances	0.954	1.00	1.00
3	Evaluating information about the use of renewable energy sources.	0.977	1.00	1.00
4	Applying the concept of energy transformation in household appliances	0.932	1.00	1.00
5	Interpreting energy usage data in the context of households	0.932	1.00	1.00
6	Evaluating information about the use of renewable energy sources.	0.954	1.00	1.00

As shown in Table 4, the content validity results for the assessment items designed to evaluate students' understanding of energy concepts demonstrate high levels of agreement among experts. The table includes three key metrics: Aiken's V, Content Validity Ratio (CVR), and Item-Level Content Validity Index (I-CVI). Aiken's V values ranged from 0.932 to 0.977, reflecting a strong consensus regarding the relevance of each item. The CVR, which evaluates the necessity of each item based on expert judgment, was 1.00 for all items, indicating unanimous agreement on their importance. Similarly, the I-CVI values were all 1.00, confirming that every expert deemed each item relevant. Collectively, these findings affirm that the assessment items are both relevant and essential for evaluating students' understanding of energy concepts, ensuring their suitability for educational purposes.

4. EFA Assumption Test

Before conducting Exploratory Factor Analysis (EFA), it is crucial to perform assumption testing to ensure the appropriateness of the data for factor analysis. Assumption testing helps to verify that the data meets the necessary conditions for EFA, such as the presence of sufficient correlations among variables, the adequacy of sample size, and the normality of the data distribution (Alavi et al., 2020; Prihono et al., 2022). This foundational step is essential for developing robust assessment instruments that accurately measure the intended constructs, such as scientific literacy skills in this study.

Table 5. EFA Assumption Test Results

Kaiser-Meyer-Olkin Test			**Bartlett's Test**			**Merdia's Test of Multivariate Normality**			
Item Indicator	**Rubric**	**MSA**	χ^2	**df**	**p**	**Value**	**Statistics**	**df**	**p**
	Overall MSA	0.864	114.914	120	0.014	Skewness=0.255	61.279	816	0.341
Explaining the concepts of potential and kinetic energy	A1	0.995				Small Sample Skewness=0.261	73.997	816	0.329
	A2	0.822							
	A3	0.816				Kurtosis=2.981	0.716		0.166
Applying the concept of energy transformation in household appliances	B1	0.859							
	B2	0.877							
Evaluating information about the use of renewable energy sources.	C1	0.838							
	C2	0.889							
	C3	0.810							
Analyzing the efficiency of energy use in household appliances	D1	0.878							
	D2	0.861							
	D3	0.873							
Interpreting energy usage data in the context of households	E1	0.877							
	E2	0.824							
Drawing conclusions from energy conversion efficiency data for the	F1	0.815							
	F2	0.827							
	F3	0.978							

Kaiser-Meyer-Olkin Test			Bartlett's Test			Merdia's Test of Multivariate Normality			
Item Indicator	Rubric	MSA	χ^2	df	p	Value	Statistics	df	p
	Overall MSA	0.864	114.914	120	0.014	Skewness=0.255	61.279	816	0.341
development of energy sources									

Table 5 presents the results of the assumptions testing for Exploratory Factor Analysis (EFA), including the Kaiser-Meyer-Olkin (KMO) Test, Bartlett's Test, and Mardia's Test of Multivariate Normality. The KMO Test helps to ensure that there are sufficient correlations among the variables, which is essential for the validity of the factor analysis results. A high KMO value suggests that the data is suitable for EFA, while a low value indicates that the analysis may not yield meaningful factors (Gibson Jr. et al., 2020; Prihono et al., 2022). Additionally, Bartlett's Test of Sphericity assesses whether the correlation matrix of the variables significantly differs from an identity matrix, indicating that the variables are correlated. A significant result ($p < 0.05$) suggests that the data is suitable for Exploratory Factor Analysis (EFA), as it confirms the presence of sufficient common variance among the variables (Alavi et al., 2020; Goretzko, 2022; Schreiber, 2021). Finally, Mardia's Test of Multivariate Normality is used to assess whether the data follows a multivariate normal distribution, which is an important assumption for Exploratory Factor Analysis (EFA). This test evaluates skewness and kurtosis in the data; significant deviations from normality can indicate that the data may not be suitable for EFA, potentially affecting the validity of the factor structure identified (Gibson Jr. et al., 2020; Karling et al., 2023; Schreiber, 2021; Siraj-Ud-Doulah, 2021).

The overall KMO value is 0.864, indicating that the data is suitable for factor analysis, as values above 0.80 are considered excellent. The individual Measure of Sampling Adequacy (MSA) values for the indicators range from 0.810 to 0.995, further supporting the appropriateness of the data for EFA. Bartlett's Test yielded a chi-square value of 114.914 with 120 degrees of freedom and a p-value of 0.014, which is below the 0.05 threshold, suggesting that the correlation matrix is significantly different from an identity matrix and confirming that there are sufficient correlations among the variables for EFA.

Additionally, Mardia's Test assessed multivariate normality, with skewness values of 0.255 and 0.261 for small samples, and kurtosis values of 2.981. The p-values for skewness (0.341) and kurtosis (0.166) indicate that the data does not significantly deviate from normality. These results collectively affirm that the data meets the necessary assumptions for conducting EFA, ensuring the validity of the subsequent analyses. The strong KMO value, significant Bartlett's Test result, and acceptable measures of normality indicate that the dataset is well-suited for exploring underlying factor structures.

5. EFA Results

Exploratory Factor Analysis (EFA) is a vital statistical technique used in research to identify the underlying relationships between measured variables and to uncover the latent constructs that may not be directly observable. By reducing data complexity, EFA helps researchers determine the number of factors that explain the correlations among variables, facilitating a deeper understanding of the data structure. This method is particularly useful in the development and validation of assessment instruments, as it allows for the refinement of measurement tools by ensuring that the items effectively capture the intended constructs (Asiye, 2022; Selau et al., 2020). Ultimately, EFA contributes to the robustness and validity of research findings, enabling more accurate interpretations and applications of the results.

Table 6. Model Fit

Chi-square test	χ^2	**df**	**p**
Model	93.198	104	0.767
Additional Fit Indices			
Comparative Fit Index (CFI)	0.938		
Tucker-Lewis Index (TLI)	0.941		
Root Mean Square Error of Approximation (RMSEA)	0.000		
RMSEA 90% Confidence Interval	0 – 0.036		
Standardized Root Mean Square Residual (SRMR)	0.052		

Table 6 presents the model fit results for the structural equation modeling analysis conducted in this study. The chi-square test yielded a value of 93.198 with 104 degrees of freedom and a p-value of 0.767. A high p-value indicates that the model fits the data well, as it suggests that there is no significant difference between the observed and expected covariance matrices. This result is favorable, as it implies that the proposed model adequately represents the relationships among the variables.

In addition to the chi-square test, several additional fit indices were evaluated to further assess the model's fit. The Comparative Fit Index (CFI) was reported at 0.938, and the Tucker-Lewis Index (TLI) was 0.941, both of which exceed the commonly accepted threshold of 0.90, indicating a good fit. The Root Mean Square Error of Approximation (RMSEA) was found to be 0.000, with a 90% confidence interval ranging from 0 to 0.036, suggesting that the model has a very low error of approximation. Finally, the Standardized Root Mean Square Residual (SRMR) was reported at 0.052, which is below the recommended cutoff of 0.08, further supporting the conclusion that the model fits the data well.

The results from the chi-square test and the additional fit indices collectively indicate that the proposed model demonstrates a strong fit to the data. These findings suggest that the model is appropriate for understanding the relationships among the constructs being studied, providing a solid foundation for further analysis and interpretation of the results.

Table 7. Factor Loadings

Item Indicator	**Assessment Rubric**	**Factor 1**	**Uniqueness**
Explaining the concepts of potential and kinetic energy	A1	0.945	0.032
	A2	1.000	0.000
	A3	0.926	0.047
Applying the concept of energy transformation in household appliances	B1	0.999	0.001
	B2	0.997	0.003
Evaluating information about the use of renewable energy sources.	C1	0.988	0.022
	C2	0.999	0.001
	C3	0.980	0.020
Analyzing the efficiency of energy use in household appliances	D1	0.808	0.048
	D2	0.988	0.012
	D3	0.973	0.029
Interpreting energy usage data in the context of households	E1	0.865	0.101
	E2	0.934	0.072
Drawing conclusions from energy conversion efficiency data for the development of energy sources	F1	0.892	0.071
	F2	0.933	0.059
	F3	1.000	0.000

Table 7 presents the factor loadings and uniqueness values for the indicators related to energy concepts, derived from the Exploratory Factor Analysis (EFA). Factor loadings indicate the strength of the relationship between each indicator and the underlying factor, with higher values suggesting a stronger association. In this analysis, all indicators demonstrate substantial loadings on Factor 1, with values ranging from 0.808 to 1.000. Notably, the indicator applying the concept of energy transformation in household appliances (B2) achieved a perfect loading of 1.000, indicating a complete association with the factor.

The uniqueness values, which reflect the proportion of variance in each indicator that is not explained by the factor, are also presented in the table. Lower uniqueness values suggest that the indicators are well-explained by the factor. For instance, the uniqueness value for indicator A2 is 0.000, indicating that it is entirely explained by Factor 1. Conversely, the indicator interpreting energy usage data in the context of households (E1) has a uniqueness value of 0.101, suggesting that a small portion of its variance is not accounted for by the factor.

The results from Table 5 indicate that the indicators are strongly associated with the underlying factor, supporting the validity of the factor structure identified in the analysis. The high factor loadings and low uniqueness values suggest that the indicators effectively measure the intended constructs related to energy concepts, providing a solid foundation for further research and application in educational contexts. This strong factor structure enhances the reliability of the assessment tools developed for evaluating students' understanding of energy concepts.

Table 8. Factor Characteristics

		Unrotated solution			**Rotated solution**		
	Eigenvalues	SumSq. Loadings	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Factor 1	1.012	0.833	0.817	0.817	0.833	0.817	0.817

Table 8 presents the characteristics of the factors identified in the Exploratory Factor Analysis (EFA), specifically focusing on the eigenvalues and the sum of squared loadings for both unrotated and rotated solutions. The eigenvalue for Factor 1 is reported as 1.012, indicating that this factor accounts for a significant amount of variance in the data. Eigenvalues greater than 1 are generally considered indicative of meaningful factors, suggesting that Factor 1 is a substantial contributor to the overall model.

In the unrotated solution, the sum of squared loadings for Factor 1 is 0.833, which represents the total variance explained by this factor. The proportion of variance attributed to Factor 1 is 0.817, indicating that it explains approximately 81.7% of the total variance in the dataset. This high proportion underscores the importance of Factor 1 in capturing the underlying structure of the data. The cumulative variance for the unrotated solution also stands at 0.817, confirming that this single factor accounts for the majority of the variance.

In the rotated solution, the sum of squared loadings remains at 0.833, with the proportion of variance and cumulative variance also reported as 0.817. The consistency of these values between the unrotated and rotated solutions suggests that the factor structure is stable and that the rotation did not significantly alter the variance explained by Factor 1. These results indicate that the identified factor effectively captures the essential characteristics of the data, providing a robust foundation for further analysis and interpretation in the context of energy concepts.

6. Reliability

In educational assessments, high reliability supports accurate evaluations of students' abilities and the validity of instructional decisions. Reliability test evaluates the internal consistency of the developed tools to ensure their reliability in measuring science literacy in the context of energy concepts.

Table 9. Reliability

Estimate	**McDoland's ω**	**Cronbach's α**
Point estimate	0.886	0.922
95% CI lower bound	0.817	0.834
95% CI Upper bound	0.956	0.911

Table 9 presents the results of the reliability analysis for the assessment tools used in this study, specifically focusing on McDonald's ω and Cronbach's α as measures of internal consistency. The point estimate for McDonald's ω is reported at 0.886, indicating a high level of reliability for the scale. Values above 0.80 are generally considered acceptable, and this result suggests that the items within the scale are consistently measuring the same underlying construct.

In addition to the point estimate, the table provides the 95% confidence interval (CI) for McDonald's ω, with a lower bound of 0.817 and an upper bound of 0.956. This range further supports the reliability of the scale, as both bounds exceed the commonly accepted threshold of 0.70, indicating that the true reliability of the scale is likely to be high. Similarly, Cronbach's α is reported at 0.922, which also reflects excellent internal consistency among the items. This value reinforces the findings from McDonald's ω, suggesting that the assessment tool is robust and reliable for measuring the intended constructs.

The results from Table 7 indicate that the assessment tools employed in this study possess strong reliability, as evidenced by both McDonald's ω and Cronbach's α. These findings provide confidence in the validity of the data collected, ensuring that the results derived from the analysis are trustworthy and reflective of the underlying constructs related to energy concepts. This high level of reliability is crucial for the subsequent interpretation and application of the findings in educational contexts.

Discussion

The findings of this study demonstrate the successful development and validation of a comprehensive essay-based assessment instrument for measuring science literacy in energy topics. Several key aspects of the results warrant further discussion:

First, the high content validity indices (Aiken's V > 0.932, CVR = 1.00, I-CVI = 1.00) across all items indicate exceptional content relevance and representation. These values surpass those reported in similar studies, such as Rusilowati et al. (2018) who achieved content validity indices ranging from 0.76 to 0.89. The strong agreement among experts suggests that the instrument effectively captures the essential aspects of science literacy in energy concepts.

Second, the EFA results revealed a robust single-factor structure explaining 81.7% of the total variance, with factor loadings ranging from 0.808 to 1.000. This finding indicates that the instrument measures a coherent construct of science literacy, aligning with theoretical frameworks that conceptualize science literacy as an integrated capability. The high factor loadings suggest that each item contributes significantly to measuring the intended construct, comparing favorably to previous studies where factor loadings typically ranged from 0.60 to 0.85 (Fathia et al., 2023).

Third, the reliability coefficients (McDonald's ω = 0.886, Cronbach's α = 0.922) demonstrate excellent internal consistency, exceeding the conventional threshold of 0.80 for high-stakes educational assessments. These values are notably higher than those reported in similar instruments, such as Hardjo et al. (2019) who reported reliability coefficients around 0.85.

The successful validation of this instrument addresses a critical gap in science education assessment by providing a reliable tool for measuring complex aspects of science literacy. The integration of analytical rubrics with essay-based assessment offers a more nuanced evaluation approach compared to traditional multiple-choice formats, enabling teachers to better understand students' conceptual understanding and critical thinking abilities in energy-related topics.

However, it is important to acknowledge that the instrument's effectiveness may vary across different educational contexts and student populations. Future research should explore its applicability in diverse settings and investigate potential modifications needed for different grade levels or cultural contexts.

CONCLUSION

This study successfully developed and validated an essay test assessment instrument aimed at measuring middle school students' science literacy skills in the context of energy. The instrument was designed based on comprehensive science literacy indicators derived from theoretical studies, curriculum analysis, and expert input. Rigorous validation processes, including the Delphi method, Aiken's V, Content Validity Ratio (CVR), and Item-Level Content Validity Index (I-CVI), confirmed the clarity, relevance, and necessity of each test item. All indicators achieved high Aiken's V values, with CVR and I-CVI values reaching 1.00, signifying unanimous agreement among experts on the importance and relevance of the items.

Quantitative analyses further reinforced the robustness of the instrument. Exploratory Factor Analysis (EFA) demonstrated strong construct validity, as evidenced by significant factor loadings and low uniqueness values, indicating that the indicators effectively measure the intended constructs. Reliability testing also yielded excellent results, with McDonald's ω and Cronbach's α values of 0.886 and 0.922, respectively, confirming very good internal consistency. These findings collectively establish that the test instrument and assessment rubric are both valid and reliable for evaluating students' science literacy skills in the context of energy.

The developed instrument provides educators with a robust tool for assessing students' understanding, critical thinking, and application of energy-related concepts in real-life contexts. Its use can help improve the evaluation of science literacy in middle school education, thereby contributing to addressing the low levels of science literacy among Indonesian students. Future research is encouraged to expand the use of this instrument in diverse educational settings and to explore its integration with innovative teaching approaches to enhance science literacy development and learning outcomes further.

REFERENCE

Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Exploratory factor analysis and principal component analysis in clinical studies: Which one should you use? *Journal of Advanced Nursing*, *76*(8), 1886–1889. https://doi.org/10.1111/jan.14377

Anggara, D. S., & Abdillah, C. (2023). Content validity analysis of literacy assessment instruments. *Jurnal Cakrawala Pendidikan*, *42*(2). https://doi.org/10.21831/cp.v42i2.55900

Asiye, Ş. A. (2022). Comparing the automatic item selection procedure and exploratory factor analysis in determining factor structure. *Participatory Educational Research*, *9*(2), 416–436. https://doi.org/10.17275/per.22.47.9.2

Colbert-Getz, J. M., Ryan, M., Hennessey, E., Lindeman, B., Pitts, B., Rutherford, K. A., Schwengel, D., Sozio, S. M., George, J., & Jung, J. (2017). Measuring assessment quality with an assessment utility rubric for medical education. *MedEdPORTAL*. https://doi.org/10.15766/mep_2374-8265.10588

Dijkstra, J., Galbraith, R., Hodges, B. D., McAvoy, P. A., McCrorie, P., Southgate, L. J., Van der Vleuten, C. P., Wass, V., & Schuwirth, L. W. (2012). Expert validation of fit-for-purpose guidelines for designing programmes of assessment. *BMC Medical Education*, *12*(1), 20. https://doi.org/10.1186/1472-6920-12-20

Fasya, N. K., Sjaifuddin, S., & Kurniasih, S. (2023). Pengembangan website pembelajaran berbasis literasi sains pada topik global warming siswa kelas VII SMP. *JURNAL PENDIDIKAN MIPA*, *13*(2), 367–374. https://doi.org/10.37630/jpm.v13i2.951

Fathia, A., Berlian, L., & Zaky El Islami, R. A. (2023). Pengembangan instrumen tes kemampuan berpikir tingkat tinggi tema energi ramah lingkungan pada siswa kelas IX. *PENDIPA*

Journal of Science Education, *7*(2), 232–240. https://doi.org/10.33369/pendipa.7.2.232-240
Gasmalla, H. E. E., & Tahir, M. E. (2021). The validity argument: Addressing the misconceptions. *Medical Teacher*, *43*(12), 1453–1455. https://doi.org/10.1080/0142159X.2020.1856802
Gibson Jr., T. O., Morrow, J. A., & Rocconi, L. M. (2020). A modernized heuristic approach to robust exploratory factor analysis. *The Quantitative Methods for Psychology*, *16*(4), 295–307. https://doi.org/10.20982/tqmp.16.4.p295
Giraldo, F., Naranjo-Trujillo, D. E., & Ariza-Villa, J. A. (2023). From the Design of Assessments to Language Assessment Literacy. *Folios*, *58*, 126–139. https://doi.org/10.17227/folios.58-16385
Goldhammer, F., Scherer, R., & Greiff, S. (2020). Editorial: Advancements in technology-based assessment: emerging item formats, test designs, and data sources. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03047
Goretzko, D. (2022). Factor retention in exploratory factor analysis with missing data. *Educational and Psychological Measurement*, *82*(3), 444–464. https://doi.org/10.1177/00131644211022031
Handriyati, D. S., Degeng, I. N. S., & Sitompul, N. C. (2022). Pengaruh strategi pembelajaran peer instruction flipped terhadap peningkatan kemampuan literasi siswa. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *7*(1), 114–120. https://doi.org/10.29100/jipi.v7i1.2482
Hardjo, F. N., Permanasari, A., & Permana, I. (2019). Meningkatkan literasi sains siswa kelas 7 melalui pembelajaran inkuiri menggunakan bahan ajar berbasis proyek pada materi energi. *JOURNAL OF SCIENCE EDUCATION AND PRACTICE*, *2*(2), 1–9. https://doi.org/10.33751/jsep.v2i2.1393
Karling, M. J., Genton, M. G., & Meintanis, S. G. (2023). *Goodness-of-fit tests for multivariate skewed distributions based on the characteristic function*.
Mardani, D. A., Farida, S. N., Supriadi, B., & Apriliyani, S. (2023). Penggunaan LKPD berbantuan simulasi phet dalam model pbl untuk meningkatkan hasil belajar kognitif siswa. *JURNAL PEMBELAJARAN FISIKA*, *12*(2), 82. https://doi.org/10.19184/jpf.v12i2.39659
Mellyzar, M., Zahara, S. R., & Alvina, S. (2022). Literasi sains dalam pembelajaran sains siswa SMP. *Pendekar: Jurnal Pendidikan Berkarakter*, *5*(2), 119. https://doi.org/10.31764/pendekar.v5i2.10097
Nikmard, F., & Mohamadi Zenouzagh, Z. (2020). Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy. *Language Testing in Asia*, *10*(1), 8. https://doi.org/10.1186/s40468-020-00106-1
OECD. (2022). *PISA 2022 Results: The State of Learning and Equity in Education: Vol. I* (Issue 2).
Prihono, E. W., Lapele, F., Jumaeda, S., Sukadari, S., & Nurjanah, S. (2022). *EFA of Pedagogic Competence Instrument to Measure Teacher Performance*. https://doi.org/10.2991/assehr.k.220129.059
Ramadani, R. D., Yennita, Y., & Ernidawati, E. (2024). Pengembangan instrumen tes berbasis literasi sains siswa smp pada materi getaran dan gelombang. *Silampari Jurnal Pendidikan Ilmu Fisika*, *6*(1), 25–34. https://doi.org/10.31540/sjpif.v6i1.2428
Rusilowati, A., Nugroho, S. E., Susilowati, E. S. M., Mustika, T., Harfiyani, N., & Prabowo, H. T. (2018). The development of scientific literacy assessment to measure student's scientific literacy skills in energy theme. *Journal of Physics: Conference Series*, *983*, 012046. https://doi.org/10.1088/1742-6596/983/1/012046
Santoso, A. N., Sunarti, T., & Wasis, W. (2023). Effectiveness of contextual phenomena-based learning to improve science literacy. *International Journal of Current Educational Research*, *2*(1), 17–26. https://doi.org/10.53621/ijocer.v2i1.205

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, *17*(5), 1004–1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Selau, T., da Silva, M. A., de Mendonça Filho, E. J., & Bandeira, D. R. (2020). Evidence of validity and reliability of the adaptive functioning scale for intellectual disability (EFA-DI). *Psicologia: Reflexão e Crítica*, *33*(1), 26. https://doi.org/10.1186/s41155-020-00164-7

Siraj-Ud-Doulah, Md. (2021). An alternative measures of moments skewness kurtosis and jb test of normality. *Journal of Statistical Theory and Applications*, *20*(2), 219. https://doi.org/10.2991/jsta.d.210525.002

Sukendar, S., & Setiawan, A. (2018). High school physics teacher's competences in designing physics lesson plan for improving student's energy literacy. *IOP Conference Series: Materials Science and Engineering*, *434*, 012016. https://doi.org/10.1088/1757-899X/434/1/012016