Development of a systems-thinking-based assessment instrument for environmental literacy and problem-solving skills in SDG 6 contexts

Leo Muhammad Taufik, Ari Widodo*, Hertien Koosbandiah Surtikanti, Taufik Rahman

Department of Science Education, Universitas Pendidikan Indonesia, Dr. Setiabudi Street No. 229, Bandung 40154, Indonesia

*Corresponding author, email: widodo@upi.edu

Article History
Received: 11 October 2025
Revised: 20 December 2025
Accepted: 2 January 2026
Published: 1 March 2026

Keywords
Environmental literacy
Instrument development
Problem solving
SDG 6
System thinking

Abstract
This study aims to develop an integrated assessment instrument that measures environmental literacy and problem-solving skills based on system thinking in the context of Sustainable Development Goal (SDG) 6: Clean Water and Sanitation. Using the Borg and Gall R&D framework, the instrument was constructed in five phases involving preliminary analysis, planning, item development, field testing, and revision. The instrument consisted of five contextual essay questions with fourteen sub-indicators aligned to environmental literacy dimensions, systems-thinking components, and Polya's problem-solving stages. The subjects were 132 tenth-grade high school students enrolled in Biology classes on the topic of Environmental Change from four different schools in Cirebon. The developed instrument consisted of five contextual essay questions with fourteen sub-questions. Content validity was analyzed using Aiken's V, construct validity using Exploratory Factor Analysis (EFA), and reliability using Cronbach's Alpha. The results showed an Aiken's V value of 0.86 (highly valid), KMO value of 0.60, and significant Bartlett's Test ($p < 0.05$). Four main factors were identified, consistent with Polya's problem-solving stages, with a Cronbach's Alpha of 0.83. The average student score was 61.4, with 2% in the high category, 35% moderate, and 63% low. These results indicate that the developed instrument is feasible to assess students' environmental literacy and problem-solving skills in the context of tenth-grade Biology learning on Environmental Change related to SDG 6. The overall low student performance suggests that learners have not yet developed a deep understanding of environmental concepts, system interactions, and structured problem-solving processes. This also reveals an instructional gap, particularly in providing opportunities for inquiry-based learning, systems thinking, and contextualized environmental analysis. Therefore, the instrument is valuable for teachers as it not only diagnoses specific weaknesses in students' competencies but also guides the refinement of instructional strategies to better integrate SDG 6 issues into biology learning.

How to cite: Taufik, L. M., Widodo, A., Surtikanti, H. K., & Rahman, T. (2026). Development of a systems-thinking-based assessment instrument for environmental literacy and problem-solving skills in SDG 6 contexts. *Journal of Environment and Sustainability Education, 4*(1). 1–11. doi: 10.62672/joease.v4i1.173

1. Introduction

Sustainable Development Goal 6 highlights one of the most urgent global sustainability challenges: ensuring access to clean water and adequate sanitation for all. Many regions continue to struggle with declining water quality, uneven distribution, and inadequate wastewater management, creating long-term risks for human health and ecosystem stability (Arora & Mishra, 2022; Gregucci et al., 2023). These challenges underscore the critical role of schools in strengthening students' awareness and understanding of water-related issues. By integrating SDG 6 into science education particularly in Biology classrooms schools can equip learners with the knowledge and skills needed to understand water systems, recognize threats to water security, and engage in responsible decision-making that supports sustainable water management.

Environmental literacy is a necessary skill for students to develop the knowledge, understanding, and responsibility required to address environmental issues. Research has shown that students often hold misconceptions about core environmental concepts, such as water cycles, pollution processes, and ecosystem interactions (Erdogan, 2020; Lin et al., 2020). Their understanding is frequently superficial, relying on memorization rather than meaningful interpretation of environmental processes. Moreover, many students demonstrate fragmented system knowledge, perceiving environmental problems as isolated events rather than interconnected phenomena within complex systems (Zimmerman & Weible, 2017; Bodin, 2017). These findings

doi: 10.62672/joease.v4i1.173

ISSN: 3025-0714

underscore the need for educational approaches and assessment tools that strengthen system-based and contextual learning.

Systems thinking provides a conceptual framework for understanding how environmental components are interconnected and how changes in one part of a system can generate consequences across the whole network. It enables students to identify relationships, feedback loops, and causal pathways that shape environmental phenomena (Eilam, 2012; Doychinova, 2023). In biology education, this perspective helps learners view environmental issues not as isolated events but as dynamic interactions among biological, physical, and social factors. However, despite its importance, systems thinking is rarely embedded in assessment practices, and few existing instruments simultaneously evaluate environmental literacy, system analysis, and contextual problem-solving skills.

Problem-solving is another essential competency in science education, as it requires students to apply scientific reasoning when addressing real environmental challenges. Polya's four-step model understanding the problem, devising a plan, carrying out the plan, and evaluating the solution provides a structured framework that guides students in making informed and logical decisions (Thuan & Giang, 2023). In the context of biology learning, students must interpret evidence, analyse causal relationships, and weigh alternative actions when dealing with issues such as waste management, water pollution, or sanitation. When combined with systems thinking, this process supports contextual decision-making, enabling students to generate solutions that are both scientifically sound and environmentally responsible (Demssie et al., 2023). Therefore, developing an instrument that integrates systems thinking, environmental literacy, and Polya-based problem-solving offers a comprehensive way to measure students' higher-order thinking skills.

Several scholars have developed assessment tools to measure students' environmental awareness or science literacy, yet many of these instruments have notable limitations. Existing tools often emphasize conceptual knowledge only, without evaluating how students apply biological concepts to environmental and social dimensions (Luft et al., 2022). They also tend to show a lack of integration between environmental literacy, systems thinking, and problem-solving, resulting in fragmented assessments that overlook the interconnected nature of sustainability issues. Moreover, many instruments are insufficiently contextualized, failing to situate tasks within meaningful real-world scenarios related to sustainable development. A further shortcoming is the heavy reliance on non-performance tasks, such as multiple-choice tests, which do not adequately capture students' reasoning, decision making, or ability to propose feasible solutions.

Given these gaps, there is a growing need for assessment tools aligned with contemporary educational priorities, particularly Education for Sustainable Development (ESD). The present study addresses this need by developing an integrated performance-based instrument that measures environmental literacy and problem-solving skills grounded in systems thinking, specifically in the context of Environmental Change for tenth-grade biology. The instrument focuses on water and sanitation issues aligned with SDG 6, linking ecological concepts to authentic contexts and encouraging students to apply biological understanding toward sustainable actions. It consists of contextual essay questions that require students to analyse, interpret, and propose solutions to real problems involving clean water, pollution, and sanitation.

The purpose of this study is to develop a comprehensive and psychometrically sound assessment instrument that measures students' environmental literacy and system-based problem-solving skills within the context of Environmental Change and SDG 6. The novelty of this work lies in its integration of systems thinking with environmental literacy and Polya-based problem solving an approach that remains scarce in Indonesian biology education, where most existing tools focus solely on conceptual understanding and lack authentic, contextualized tasks. By offering a performance-based instrument grounded in contemporary sustainability demands, this study provides both theoretical and practical contributions for strengthening curriculum design and classroom assessment practices. To guide this work, the study addresses the following research questions:

a. How can an integrated assessment instrument be developed to accurately measure students' environmental literacy and problem-solving skills using systems thinking in the context of Environmental Change?

b. What are the psychometric properties (content validity, construct validity, and reliability) of the developed instrument?

c. How are students' environmental literacy and problem-solving skills distributed in relation to SDG 6: Clean Water and Sanitation?

These research questions shaped the design, validation, and interpretation of the instrument, ensuring that the final product meets scientific standards while offering meaningful contributions to sustainable biology education.

2. Method

2.1. Research Design

This study used Borg and Gall's (1983) Research and Development (R&D) methodology, which was adapted for educational assessment development. The model was selected because it systematically guides product creation through iterative expert validation and field testing, a process essential for developing psychometrically robust educational instruments. Five stages were implemented in this study: (1) preliminary research and information gathering, (2) planning and framework design, (3) initial product development, (4) limited field testing, and (5) product revision. Each stage ensured that the resulting instrument demonstrated both content and construct validity and was suitable for use in high school biology education. The methodological procedure was intended to be reproducible, allowing future researchers to replicate or alter the development processes in similar educational situations.

2.2. Participants

The participants in this study were 132 tenth-grade high school students drawn from four senior high schools located in Cirebon, West Java, Indonesia. The sample consisted of two schools from Cirebon Regency and two schools from Cirebon City, representing diverse geographical and socio-economic characteristics within the region. This distribution ensured that the developed instrument could capture variations in students' understanding across urban and rural contexts. The participants were enrolled in Biology classes, specifically studying the topic of Environmental Change, which was aligned with the learning objectives of SDG 6: Clean Water and Sanitation. Prior to the research implementation, permission was obtained from school principals, and ethical approval was secured. All participants were informed about the purpose of the study and voluntarily agreed to participate.

2.3. Instrument Development

The developed assessment instrument consisted of five contextual essay questions comprising fourteen sub-questions. Each item was constructed to measure three main constructs: (1) environmental literacy, (2) system thinking, and (3) problem-solving skills based on Polya's four-stage model (understanding, planning, implementing, and evaluating). The content was drawn from the national Biology curriculum (Grade X) and contextualized with real-world issues such as water pollution, sanitation management, and waste disposal. The initial draft was reviewed by a panel of three experts two science educators and one assessment specialist who evaluated the items for relevance, clarity, and integration with SDG 6 themes. Accordingly, the developed instrument was structured by mapping the indicators of environmental literacy, system thinking, and problem-solving skills into contextual essay questions. Each indicator was associated with a specific item and categorized into levels of student responses ranging from low, moderate, to high understanding. The distribution of indicators and response levels is presented in Table 1.

Table 1. Distribution of Indicators and Response Levels

Question	Indicator	High Level	Medium Level	Low level
Q1a – Mention 2 problems if Cirebon takes water from the river	EL1 (Essential Water and Impacts) PS1 (Problem Identification)	Mention ≥2 relevant issues (pollution, decreasing discharge, conflict)	Mention 1-2 common problems (dirty water, low flow)	Calling it very simple ("*dirty water*", "*sick person*")
Q1b – Long-term impact	EL1 (Essential Water and Impacts) ST2 (Cause and Effect) ST6 (Systemic Solutions) PS1 (Problem Analysis)	Complete impact on health, environment, rivers	Limited impact on health & environment	Short/general impact, only "*sick society*"
Q2a – Diagram of rain, rivers, groundwater, people	EL2 (Water in the Global System) ST1 (Components), ST2 (Cause and Effect) ST3 (System Visualization) ST5 (Feedback)	Clear diagram with arrows (+/-), there are stars on key factors	Simple diagram, only 2-3 relationships	Basic relationship, without notation/feedback
Q2b – Problem if 1 part is disturbed	EL2 (Water in the global system) ST2 (Cause-Effect) PS1 (Problem Identification)	Mention ≥2 relevant issues (drought, crop failure, pollution)	Mention 1 common problem	Brief mention, without details

Question	Indicator	High Level	Medium Level	Low level
Q2c – Impact of reduced groundwater	EL2 (Water in the global system) ST2 (Cause-Effect)	Mentioning the impacts on rivers and humans	Name one (river or human)	Short/inappropriate answer
Q3a – Influence of water quality, sanitation, habits	EL1 (Essential Water & Impact) EL3 (Access & Sanitation) ST2 (Cause and Effect) PS1 (Problem Analysis)	Explain all three in full, with examples.	Only 1-2 aspects, for example water & sanitation	Brief, unsystematic mention
Q3b – Two realistic solutions	EL3 (Access and Sanitation) PS3 (Alternative Solutions)	≥2 realistic solutions (pipes, infiltration wells, education)	1 common solution (pipe/filter)	Instant solution ("*give water*")
Q3c – Best solution + reasons	EL3 (Access and Sanitation) ST6 (Systemic Solutions) PS4 (Evaluation of Best Solutions) PS5 (Implementation & Reflection)	Have 1 solution, give reasons and how to implement it	Choosing a solution but with minimal reasons	Calling solutions without reason
Q4a – Calculate the total virtual water	EL4 (Virtual water)	Correct Count (8 L)	Count correctly	Count correctly
Q4b – The most wasteful stage	EL4 (Virtual Water) PS1 (Problem Analysis)	Identify "wash the pot" & clear reasons	Call it a wasteful stage but the reason is simple	Calling without reason
Q4c – Two solutions + impact	EL4 (Virtual Water) ST6 (Systemic Solution) ST7 (Impact of the Solution) PS3 (Alternative Solution) PS5 (Reflection)	≥2 solutions and long-term impacts (rivers, costs, environment)	1 solution with short impact	General solution without long term impact
Q5a – Major river problems	EL5 (IWRM: flood, drought) PS1 (Problem Identification)	Mentioning all (floods, drought, garbage)	Mentioning some	Just 1 common problem
Q5b – River diagram	EL5 (IWRM: floods, droughts) ST1 (Components), ST2 (Cause-Effect) ST3 (System Visualization) ST4 (Key Components) ST5 (Feedback)	Complete diagram with (+/-) & star notation	Simple diagram, lack of feedback	Minimal/incorrect diagram
Q5c – Cooperation & 10-year projections	EL5 (IWRM: floods, droughts) ST7 (Solution Impact) PS5 (Implementation & Reflection)	Explain the roles of government, schools, citizens and clear projections	Mentioning cooperation, short projections	General solution, unclear projection

Description: EL = environmental literacy, PS = problem solving, and ST = systems thinking.

2.4. Validation Procedures

Two types of validation were conducted in this study, namely content validity and construct validity. Content validity was evaluated using Aiken's V index based on expert judgments, which examined four aspects: the alignment of indicators with the intended constructs, the integration of environmental literacy and systems thinking, the clarity of language used in the items, and the appropriateness of item difficulty. Construct validity, on the other hand, was examined through Exploratory Factor Analysis (EFA) using principal component extraction with varimax rotation. Prior to conducting EFA, the Kaiser Meyer Olkin (KMO) test and Bartlett's Test of Sphericity were performed to assess sampling adequacy and data suitability for factor analysis. The analysis followed established threshold criteria, with KMO values greater than 0.50 and a Bartlett's test significance level of $p < 0.05$, indicating that factor extraction was statistically justified.

2.5. Reliability Testing

Cronbach's Alpha (α) was used for reliability analysis, assessing internal consistency across all test items. A coefficient value of 0.70 was deemed satisfactory, whereas values above 0.80 indicated strong reliability (Tavakol & Dennick, 2011). The reliability test confirmed that the created instrument measured the intended constructs consistently across the sample group.

2.6. Data Analysis Techniques

Descriptive and inferential statistical analyses were performed with IBM SPSS software. Means, standard deviations, and frequency distributions were used to describe student performance in three categories (high, moderate, and low). The inferential analysis included Exploratory Factor Analysis (EFA) for construct validation because the underlying factor structure of the newly developed instrument had not been previously established, making EFA more appropriate than Confirmatory Factor Analysis (CFA) for identifying emerging dimensions. Cronbach's Alpha was used to assess reliability, as it is suitable for evaluating internal consistency across open-ended performance tasks that are scored using standardized analytic rubrics. All statistical tests had a level of significance of $p < 0.05$. The distribution of students' ability levels was given as a percentage to show overall trends in environmental literacy and problem-solving performance.

2.7. Scoring Procedures and Inter-Rater Reliability

Given that the assessment instrument consists of contextual essay questions, a transparent and rigorous scoring procedure was implemented. First, a detailed analytic scoring rubric was developed to evaluate responses across the three constructs environmental literacy, systems thinking, and problem-solving skills. The rubric was reviewed by three experts in science education and assessment to ensure clarity, relevance, and alignment with the intended indicators. This process established the rubric's content validity.

To ensure scoring consistency, two independent raters were trained using exemplar responses representing high, medium, and low level performance for each indicator. The raters practiced scoring a sample of student responses and discussed discrepancies to reach a shared interpretation of the scoring criteria.

Inter-rater reliability (IRR) was calculated using Cohen's Kappa to evaluate scoring agreement between the raters. A Kappa value above 0.75 was considered acceptable, indicating substantial to excellent agreement. Only after satisfactory IRR was achieved were the remaining student responses scored. This procedure ensured that the essay-based instrument maintained scoring objectivity and reliability.

3. Results and Discussion

3.1. Results

The results of this study present the outcomes of content validity testing, construct validity analysis, reliability measurement, and students' performance distributions. All findings are presented concisely using tables and relevant statistical indicators.

3.1.1. Content Validity

Content validity was assessed using Aiken's V based on expert ratings. Table 2 summarizes the mean scores and validity coefficients for each aspect. All Aiken's V values ranged from 0.83 to 0.89, exceeding the minimum criterion of 0.75 for strong relevance. These values indicate adequate alignment between indicators and constructs, clarity of item wording, and appropriate difficulty levels.

Table 2. Content Validity Analysis (Aiken's V)

Aspect	Mean Score	Aiken's V	Category
Alignment of indicators with constructs	4.33	0.86	Highly valid
Integration of environmental literacy, system thinking, and problem-solving	4.33	0.86	Highly valid
Clarity of language and SDG context	4.67	0.89	Highly valid
Appropriateness of difficulty level	4.00	0.83	Valid
Overall Average	4.33	0.86	Highly valid

3.1.2. Construct Validity

To verify the underlying dimensional structure of the instrument, Exploratory Factor Analysis (EFA) was conducted using Principal Component Analysis (PCA) with Varimax rotation in IBM SPSS Statistics 26. Sampling adequacy met minimal requirements, with a KMO value of 0.60 and a significant Bartlett's Test of Sphericity ($p < 0.001$), indicating that the data were suitable for factor extraction.

Eigenvalues greater than 1.00 were used to determine the number of factors retained, following Kaiser's criterion. Four factors exceeded this threshold, with eigenvalues of 3.42, 2.15, 1.54, and 1.08, respectively. Together, these factors explained 62.7% of the total variance, which is considered acceptable for instruments measuring educational constructs.

A rotated factor matrix was generated to examine item loadings on each component. Most items loaded strongly (> 0.50) on a single factor, indicating clear factor structure, while no substantial cross-loadings were observed. A summarized description of the factor distribution is provided here:

a. Factor 1: Items related to understanding environmental problems

b. Factor 2: Items reflecting planning and reasoning

c. Factor 3: Items associated with implementing solutions and system interactions

d. Factor 4: Items reflecting evaluation, reflection, and impact analysis

The rationale for retaining exactly four factors was based on three criteria:

a. eigenvalues > 1.00,

b. the clear break in the scree plot after the fourth factor, and

c. theoretical alignment with Polya's four-stage problem-solving framework and systems-thinking components.

Taken together, these results confirm that the instrument possesses a coherent and theoretically grounded factor structure suitable for further reliability assessment. Construct validity test (EFA results) can be seen in Table 3.

Table 3. Construct Validity Test (EFA Results)

Statistical Test	Value	Criteria	Description
KMO	0.60	> 0.5	Acceptable for factor analysis
Bartlett's Test of Sphericity (Sig.)	< 0.001	< 0.05	Significant
Number of extracted factors	4	Consistent with Polya's theory	Valid

3.1.3. Reliability Analysis

To examine the internal consistency of the developed instrument, a reliability analysis was conducted using the Cronbach's Alpha (α) coefficient in IBM SPSS Statistics 26. Data from 132 student responses were entered and item total correlations were computed through the Reliability Analysis procedure. Each construct environmental literacy, systems thinking, and problem-solving was tested individually, and an overall reliability coefficient was also calculated. Cronbach's Alpha values exceeded the recommended threshold of 0.70, indicating satisfactory to strong internal consistency (Tavakol & Dennick, 2011; George & Mallery, 2024; Hair, 2014). Reliability statistics (Cronbach's Alpha Results) can be seen in Table 4.

Table 4. Reliability Statistics (Cronbach's Alpha Results)

Construct	Number of Items	Cronbach's Alpha (α)	Interpretation
Environmental Literacy	5	0.79	Acceptable
System Thinking	4	0.82	High reliability
Problem-Solving Skills	5	0.85	High reliability
Overall Instrument	14	0.84	High internal consistency

3.1.4. Students' Performance Analysis

To evaluate students' overall performance based on the developed instrument, descriptive statistical analysis was carried out using IBM SPSS Statistics 26. Each student's total score was calculated from the rubric-based assessment of 14 sub-items representing the constructs of environmental literacy, system thinking, and problem-solving skills. The scores were converted into percentage ranges and categorized into three levels of achievement: high (≥ 80%), moderate (60–79%), and low (< 60%). Frequency and percentage distributions were computed to identify the general pattern of students' abilities. The results of this analysis are summarized in Table 5.

Table 5. Distribution of Students' Performance Levels

Performance Level	Frequency (n)	Percentage (%)	Description
High	3	2%	Demonstrates comprehensive understanding, clear system reasoning, and accurate contextual solutions.
Moderate	46	35%	Shows adequate understanding and partial application of system-based reasoning with some conceptual gaps.
Low	83	63%	Demonstrates limited conceptual understanding and difficulty applying system thinking to problem contexts.
Total	132	100%	—

3.2. Discussion

3.2.1. Interpretation of Psychometric Findings

The results presented in Table 2 show that all aspects of the instrument achieved high levels of content validity, with Aiken's V coefficients ranging from 0.83 to 0.89, exceeding the minimum threshold of 0.75 recommended for strong item relevance (Ismarau Tajuddin et al., 2025; Ventura-León et al., 2025). The mean expert rating across all items was 4.33 out of 5, indicating strong agreement among validators regarding the quality and representativeness of each question. The highest validity value (0.89) was obtained for the aspect of language clarity and contextual relevance, confirming that the items were clearly written and directly aligned with SDG 6 themes, clean water and sanitation. This suggests that the instrument successfully integrates cognitive, contextual, and sustainability-oriented elements within Biology learning.

The high content validity scores indicate that experts consistently judged the items to be relevant and aligned with the constructs measured. Strong reliability coefficients demonstrate that the instrument provides stable and internally consistent measurements. The four-factor structure emerging from EFA corresponds with the theoretical foundations of environmental literacy, systems thinking, and problem-solving, confirming the instrument's multidimensionality.

These findings are consistent with previous studies emphasizing that clearly defined indicators and contextual relevance contribute significantly to the construct coherence of assessment instruments (Clark & Watson, 2019; Vogt et al., 2013; Ventura-León et al., 2025). The strong content validity indicates that the developed items appropriately capture students' conceptual understanding, reasoning ability, and system-based thinking. Consequently, the validated instrument can be confidently used for further testing of construct validity and reliability in subsequent stages of analysis.

3.2.2. Comparison with Previous Studies

Previous validation studies utilising EFA for educational instruments produced similar results, with moderate KMO values (0.60-0.70) deemed acceptable when the theoretical underpinning firmly supported factor interpretation (Costello & Osborne, 2005; Taber, 2018). Thus, the findings of this study indicate that the instrument has a well-defined internal structure that is consistent with the conceptual frameworks of environmental literacy and system-based problem resolution. As a result, the factor solution provided validates the instrument's multidimensionality and serves as a solid platform for future reliability study. These results align with prior research showing the importance of contextual indicators and systems-based constructs in producing coherent assessment tools. The moderate KMO values are comparable to similar validation studies in science education, supporting the adequacy of the factor solution.

3.2.3. Interpretation of Student Performance

As shown in Table 4, the overall Cronbach's Alpha value of 0.84 indicates a high degree of internal consistency, confirming that all items function coherently to measure the intended constructs. Each subscale also met or exceeded the 0.70 threshold, suggesting that the items within each construct are homogeneous and measure a unified concept. The reliability coefficients fall within the range considered satisfactory to excellent for educational assessment instruments (Taber, 2018; George & Mallery, 2024).

These findings demonstrate that the developed instrument is statistically stable and suitable for further application in classroom assessment and research. Similar studies on science literacy and environmental education instruments have also reported Cronbach's Alpha values between 0.80 and 0.90, which are considered strong indicators of measurement consistency (Kline, 2013; Hair, 2014). Therefore, the high internal reliability obtained in this study supports the quality of the instrument as a valid and dependable tool for assessing students' environmental literacy and problem-solving ability based on system thinking.

The predominance of low and moderate performance suggests that many students struggle to integrate environmental concepts with system reasoning and applied problem solving. This echoes findings from previous studies reporting fragmented system understanding among high school learners (Demssie et al., 2023; OECD, 2018). The small proportion of high performers indicates limited mastery of SDG-related reasoning skills.

The high reliability coefficients can be attributed to three key factors: (1) well defined constructs, which ensured that items within each domain measured a clearly bounded conceptual area; (2) strong alignment between indicators and their intended theoretical dimensions, resulting in items that consistently tapped into the same underlying skills; and (3) adequate item consistency, reflected in strong item total correlations observed across the instrument. These conditions collectively contributed to the robustness of internal consistency, confirming that the developed instrument reliably measures students' environmental literacy, systems thinking, and problem solving skills.

3.2.4. Educational Implications

The data in Table 5 show that a majority of students (63%) fell into the low performance category, while 35% demonstrated moderate ability, and only 2% achieved high levels of performance. The mean total score of 64.5 (SD=8.3) indicates that most students possessed only basic to intermediate understanding of environmental literacy and problem-solving within the SDG 6 context. These results suggest that while students are able to identify surface-level environmental issues, they often struggle to connect causal relationships among system components or to propose scientifically grounded solutions.

A deeper analysis reveals specific weaknesses across each construct. In terms of environmental literacy, many students found it difficult to explain long-term impacts of water pollution or sanitation problems, often giving general or descriptive statements rather than analytical explanations. For systems thinking, the weaknesses were more pronounced: students frequently treated environmental issues as isolated problems and showed limited ability to identify feedback loops, key components, or cross-scale interactions. This aligns with previous studies showing fragmented conceptions of environmental systems among secondary students (Demssie et al., 2023; Molderez & Fonseca, 2018; OECD, 2018). In the domain of problem-solving, most students struggled with later stages of Polya's model especially evaluating solutions and providing implementation steps indicating insufficient practice in scientific reasoning and contextual decision-making.

These performance patterns point to important educational implications. The dominance of low and moderate categories suggests that current instructional practices may not yet integrate systems thinking or inquiry-based approaches effectively. Strengthening the use of investigative tasks, real-world case studies, and problem-based scenarios could help students make deeper connections between biological concepts and sustainability challenges. Prior work indicates that students' performance in sustainability-related contexts improves when they experience inquiry-based and project-based learning environments that emphasize reflection, reasoning, and system-level analysis (Costa et al., 2024; Nagarajan & Overton, 2019).

In the context of SDG 6, these findings highlight an urgent need for classroom innovations that explicitly bridge environmental literacy with systems reasoning and contextual problem-solving. The assessment tool developed in this study can therefore serve not only as an evaluative instrument but also as a formative diagnostic tool that helps teachers identify specific weaknesses in students' understanding. By doing so, it supports pedagogical reforms aligned with Education for Sustainable Development (ESD), encouraging teachers to create learning experiences that promote higher-order thinking, sustainability-oriented reasoning, and evidence-based decision-making. Thus, the instrument offers practical and theoretical value for improving how environmental change and SDG 6 topics are taught and assessed in biology classrooms.

3.2.5. Theoretical Contributions

This study contributes a validated multidimensional framework that integrates environmental literacy, systems thinking, and Polya-based problem solving, addressing long-standing gaps in existing assessment tools particularly in Indonesian biology education. By combining these constructs within a single instrument, the study advances theoretical understanding of how students engage with complex environmental issues that require simultaneous conceptual, analytical, and procedural reasoning. The framework also extends current models of assessment by operationalizing systems thinking indicators and contextualizing problem-solving processes within SDG-related scenarios, an approach that is still rarely implemented in science education research. Moreover, the integration of sustainability contexts strengthens the alignment between cognitive theories of learning and contemporary ESD principles, demonstrating how theoretical constructs can be translated into measurable indicators that reflect real-world environmental decision-making. As such, the study offers a conceptual foundation that future researchers can adopt or adapt in developing assessments for other sustainability themes or interdisciplinary science domains.

3.2.6. Practical Recommendations

Schools and teachers can adopt this instrument as both an evaluative and formative tool to gain a clearer picture of students' strengths and weaknesses in environmental literacy, systems thinking, and problem solving skills. Its contextualized nature supports classroom practices aligned with SDG 6 by enabling teachers to diagnose misconceptions, identify gaps in system reasoning, and monitor students' progress as they engage with water and sanitation related learning activities. The instrument can also serve as a guide for designing instructional interventions such as inquiry-based projects, system-mapping exercises, community-based investigations, or case-study analyses that strengthen students' ability to connect biological concepts with real-world sustainability challenges. Furthermore, its use encourages schools to integrate ESD-oriented competencies more systematically into their science curriculum, promoting deeper reflection, evidence based decision making, and environmentally responsible behavior among students. In this way, the instrument provides a practical roadmap for teachers seeking to enhance both cognitive outcomes and sustainability-focused learning experiences in Biology classrooms.

3.2.7. Limitations and Future Research

This study is limited by its sample size and geographical focus, as the participants were drawn from four senior high schools within the Cirebon region. While the sample provides meaningful insights into urban rural variations, the findings may not fully represent the broader diversity of Indonesian secondary students. Future research should therefore expand the participant pool across multiple provinces or school types to enhance generalizability. Additionally, the present study captures student performance at a single point in time, which constrains interpretation of developmental changes in environmental literacy, systems thinking, and problem solving skills. Longitudinal studies would be valuable for examining how these competencies evolve through sustained instructional interventions or curriculum reforms. Further research could also explore the integration of this instrument into digital assessment platforms, enabling automated scoring, richer analytics, and more dynamic representations of system thinking tasks. Finally, investigating how the instrument performs in different pedagogical environments such as project based learning, socio scientific issue instruction, or community based sustainability projects would offer deeper insights into its adaptability and instructional impact.

4. Conclusion

The development and validation of this integrated assessment instrument provide a conceptual synthesis of how environmental literacy, systems thinking, and Polya-based problem solving can be operationalized within the context of SDG 6. Beyond demonstrating strong psychometric properties, the study shows that these constructs can be meaningfully combined to capture the multidimensional nature of sustainability-oriented reasoning in Biology learning. The findings reinforce the theoretical contribution of this work: offering a coherent framework that links cognitive understanding, system-based analysis, and contextual decision making, which remains underdeveloped in current science education assessment models.

Practically, the instrument offers teachers a dual function. It serves as a robust evaluative tool for measuring students' higher-order thinking skills, while also providing formative feedback that helps diagnose specific weaknesses in conceptual understanding, system reasoning, and solution evaluation. Its contextualized design supports classroom practices aligned with SDG 6 and encourages the adoption of inquiry-based, project-based, and socio-scientific learning approaches that foster deeper engagement with environmental issues.

This study does, however, have limitations, particularly in its regional scope and cross-sectional design. The sample drawn from Cirebon may not represent broader student populations, and a single measurement point limits insights into growth over time. Future research should therefore expand the geographic coverage, examine longitudinal development of sustainability competencies, and explore the instrument's application across diverse subjects and grade levels. Further studies could also investigate digital or adaptive assessment formats that allow richer system-mapping tasks and automated feedback, strengthening the role of assessment in advancing environmental and sustainability education.

Author Contributions

All authors have equal contributions to the paper. All the authors have read and approved the final manuscript.

Funding

No funding support was received.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The datasets generated during and/ or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration on AI Use

The authors declare that no artificial intelligence (AI) or AI-assisted tools were used in the preparation of this manuscript. AI were used only to improve readability and language under strict human oversight; no content, ideas, analyses, or conclusions were generated by AI.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. https://doi.org/10.1177/0013164485451012

Arora, N. K., & Mishra, I. (2022). Sustainable Development Goal 6: Global water security. *Environmental Sustainability*, 5(3), 271–275. https://doi.org/10.1007/s42398-022-00243-2

Bodin, Ö. (2017). Collaborative environmental governance: Achieving collective action in social-ecological systems. *Science,* 357(6352), eaan1114. https://doi.org/10.1126/science.aan1114

Borg, W. R., Gall, M. D., & Gall, M. D. (1983). *Educational research* (Rev. ed.). Longman.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427. https://doi.org/10.1037/pas0000626

Costa, J., Alscher, P., & Thums, K. (2024). Global competences and education for sustainable development: A bibliometric analysis to situate the OECD global competences in the scientific discourse. *Zeitschrift für Erziehungswissenschaft*, 1–20. https://doi.org/10.1007/s11618-023-01298-0

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 1–9.

Demssie, Y. N., Biemans, H. J., Wesselink, R., & Mulder, M. (2023). Fostering students' systems thinking competence for sustainability by using multiple real-world learning approaches. *Environmental Education Research*, 29(2), 261–286. https://doi.org/10.1080/13504622.2022.2079198

Doychinova, K. (2023). Teaching methods based on constructivism in environmental education. *Acta Scientifica Naturalis*, 10(2), 34–44.

Eilam, B. (2012). System thinking and feeding relations: Learning with a live ecosystem model. *Instructional S*cience, 40(2), 213–239. https://doi.org/10.1007/s11251-011-9179-4

Erdoğan, Ş. (2020). Microplastic pollution in freshwater ecosystems: A case study from Turkey. *Ege Journal of Fisheries and Aquatic Sciences, 37*(3), 213-221. https://doi.org/10.12714/egejfas.37.3.02

Fang, W. T., Hassan, A. A., & LePage, B. A. (2022). Environmental literacy. In The living environmental education: Sound science toward a cleaner, safer, and healthier future (pp. 93–126). *Springer Nature Singapore*. https://doi.org/10.1007/978-981-16-8500-3_6

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). Sage.

George, D., & Mallery, P. (2024). IBM SPSS statistics 29 step by step: A simple guide and reference. *Routledge*.

Gregucci, D., Nazir, F., Calabretta, M. M., & Michelini, E. (2023). Illuminating progress: The contribution of bioluminescence to Sustainable Development Goal 6 Clean water and sanitation. *Sensors*, 23(16), 7244. https://doi.org/10.3390/s23167244

Hair, J. F. (2014). A primer on partial least squares structural equation modeling (PLS-SEM). *Sage*.

Ismarau Tajuddin, N. I., Abas, U. H., Aziz, K. A., Haizan Nor, R. N., Izni, N. A., Sudin, M. N., ... & Noor, N. M. (2025). Content validity assessment using Aiken's V: Knowledge Integration Model for blockchain in higher learning institutions. *International Journal of Advanced Computer Science & Applications*, 16(6), 1–9.

Kline, P. (2013). Handbook of psychological testing (2nd ed.). *Routledge*.

Lin, Y.-L., Wu, L.-Y., Tsai, L.-T., & Chang, C.-C. (2020). The Beginning of Marine Sustainability: Preliminary Results of Measuring Students' Marine Knowledge and Ocean Literacy. *Sustainability, 12*(17), 7115. https://doi.org/10.3390/su12177115

Luft, J. A., Jeong, S., Idsardi, R., & Gardner, G. (2022). Literature reviews, theoretical frameworks, and conceptual frameworks: An introduction for new biology education researchers. *CBE Life Sciences Education*, 21(3), rm33. https://doi.org/10.1187/cbe.21-11-0270

Molderez, I., & Fonseca, E. (2018). The efficacy of real-world experiences and service learning for fostering competences for sustainable development in higher education. *Journal of Cleaner Production*, 172, 4397–4410. https://doi.org/10.1016/j.jclepro.2017.04.062

Nagarajan, S., & Overton, T. (2019). Promoting systems thinking using project- and problem-based learning. *Journal of Chemical Education*, 96(12), 2901–2909. https://doi.org/10.1021/acs.jchemed.9b00358

OECD. (2019). PISA 2018 assessment and analytical framework. *OECD Publishing*. https://doi.org/10.1787/b25efab8-en

Polya, G. (1957). How to solve it (2nd ed.). *Princeton University Press*.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thuan, H. T., & Giang, T. T. H. (2023). Application of G. Polya's problem-solving process in teaching high-school physics. *Journal La Sociale*, 4(1), 26–33. https://doi.org/10.37899/journallasociale.v4i1.1377

Ventura-León, J., Lluncor-Vasquez, A., & Chuquillanqui-Silva, M. (2025). A computer tool for Aiken's coefficients calculation: Applications in clinical and research settings. *Journal of Evaluation in Clinical Practice*, 31(2), e14134. https://doi.org/10.1111/jep.14134

Vogt, K., Jenny, G. J., & Bauer, G. F. (2013). Comprehensibility, manageability, and meaningfulness at work: Construct validity of a scale measuring work-related sense of coherence. *SA Journal of Industrial Psychology*, 39(1), 1–8. https://doi.org/10.4102/sajip.v39i1.1118

Zimmerman, H. T., & Weible, J. L. (2017). Learning in and about rural places: Connections and tensions between students' everyday experiences and environmental quality issues in their community. *Cultural Studies of Science Education*, 12(1), 7–31. https://doi.org/10.1007/s11422-015-9683-4