IJOLAE Indonesian Journal on Learning and Advanced Education

http://journals.ums.ac.id/index.php/ijolae

Automatic Assessment-Based Artificial Intellegent to Measure Students Environmental Literacy

Fitri April Yanti[1]✉, Rendy Wikrama Wardana[2], Buyung[3], Elza Heryensi[4], Noorzana Khamis[5]
[1-4]Faculty of Teacher Training and Education, Universitas Bengkulu, Indonesia
[5]Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia, Malaysia

DOI: 10.23917/ijolae.v7i3.11240
Received: June 17th, 2025. Revised: July 20th, 2025. Accepted: August 3rd, 2025
Available Online: September 1st, 2025. Published Regularly: September, 2025

Abstract

There is a need to measure the environmental literacy of secondary school students to promote awareness of environmental issues and provide sustainable solutions. Environmental literacy assessment is done using the UNESCO Criteria on Education and Culture Sustainable Development (ECSD). This study aimed to develop an automatic evaluation based on Artificial Intelligence (AABAI) to measure students' environmental literacy in science subjects and environmental issues topics. This study uses the ADDIE development model, which includes analysis, design, development, implementation, and evaluation stages. The data were examined through both quantitative and qualitative approaches using Item Response Theory (IRT). Content validity of the AABAI instrument was confirmed through Aiken's V analysis, indicating acceptable validity levels. All test items were deemed valid based on the Rasch model, considering item fit and difficulty parameters. The AABAI instrument has been implemented in educational settings and has successfully provided a comprehensive overview of junior high school students' environmental literacy profiles. The use of AABAI as a transformative assessment can ensure that students care more about the environment and maintain environmental sustainability.

Keywords: AI-based assessment, automatic assessment-based artificial intelligence, contextual learning, critical thinking skills, cultural sustainability, environmental literacy, sustainable development goals

Corresponding Author:
Fitri April Yanti, Faculty of Teacher Training and Education, Universitas Bengkulu, Indonesia
Email: faprilyanti@unib.ac.id

1. Introduction

The global plastic waste crisis is a pressing environmental issue, with serious implications for ecosystems, public health, and coastal economies. Each year, an estimated 8 million tons of plastic waste enter the oceans, severely damaging marine life and biodiversity (Jambeck et al., 2015; Aliviyanti et al., 2022; Law et al., 2020). As an archipelagic nation with the second-longest coastline globally, Indonesia plays a significant role in this crisis, contributing approximately 3.2 million tons of plastic waste annually. This problem is particularly acute in coastal provinces, including

Bengkulu, where waste accumulation has been observed to degrade seawater quality, coral reefs, and mangrove habitats ecosystems crucial to both biodiversity and coastal protection (Masrobi et al., 2022). Beyond ecological damage, this environmental degradation threatens the cultural sustainability of local communities in Bengkulu. For generations, coastal communities have depended on the sea not only for food and livelihood but also as a source of cultural identity and inspiration (Hohn et al., 2013). However, the infiltration of modern waste especially plastic into daily life increasingly disrupts these traditions, weakening the intergenerational transmission of local ecological wisdom (Afriansyah & Sukmayadi, 2022; Akbar & Maghfira, 2023; Dwiyanti Suryono, 2019). Yet, limited educational interventions address this linkage between environmental degradation and cultural erosion, especially in formal science education (Samidjo et al., 2023).

In response to this challenge, environmental literacy has emerged as a key educational goal. It encompasses not only knowledge of environmental issues but also the skills, attitudes, and behaviors necessary for responsible ecological decision-making (Hungerford & Peyton, 1976). Strengthening environmental literacy in schools is essential to building future generations capable of understanding the complex interactions between human activity and nature, especially in vulnerable regions like coastal Bengkulu.

Science education, therefore, holds a strategic position in fostering environmental literacy. Globally, there is growing attention to environmental literacy in the context of sustainability education, as shown in recent bibliometric studies (Ariyatun et al., 2024). In Indonesia, curricular integration of environmental themes has been promoted, yet assessments of students' environmental literacy often rely on manual scoring, which is time-consuming, prone to subjectivity, and lacks contextual sensitivity. Moreover, few instruments integrate local cultural relevance, especially coastal ecological knowledge, into assessments limiting their ability to connect students' lived experiences with scientific learning.

Recent research highlights the need for innovative, technology-driven assessments. AI-based platforms can offer personalized, efficient, and culturally contextualized evaluations. For instance, spirituality-based instruments have shown promise in linking values with sustainability behavior (Husamah et al., 2023; Szczytko et al., 2019), yet they remain manual in nature. To address this gap, a more advanced system is needed.

This study introduces the Automatic Assessment-Based Artificial Intelligence (AABAI) system, designed to measure students' environmental literacy through contextualized, culturally relevant, and automated assessments. AABAI specifically targets environmental issues in coastal contexts, integrating multimedia elements and AI algorithms to personalize the assessment experience. It also provides immediate, detailed feedback, enabling students to better understand environmental challenges and solutions, especially those impacting local ecosystems and culture.

This assessment measures students' theoretical knowledge and their understanding and response to environmental issues around them (Memarian & Doleck, 2024; Widarmanto, 2018). This research focuses on environmental issues related to coastal areas. Direct and automatic feedback can help students better understand the impact of waste on the environment and give them practical insights on how to reduce pollution on the coast (González-Calatayud et al., 2021; Hamid et al., 2022; Handiyati et al., 2023; Wang et al., 2024). This approach also strongly supports the achievement of the Sustainable Development Goals (SDGs), especially SDG 4 (Quality Education) and SDG 14 (Marine Life) (Aryanti et al., 2017; Putra & Dupuy, 2023; Zhai et al., 2021). Using AI in assessment can ensure that the education provided teaches basic knowledge and forms attitudes and actions that support environmental conservation (Krstić et al., 2022; Puspita, 2017). AABAI, as a transformative assessment, can help ensure that students are more concerned about the environment and preserving culture.

Through AABAI, the environmental literacy of junior high school students will be measured. AABAI will integrate materials with contextual problems in the coastal student environment, such as environmental management practices, utilization of natural resources, and local traditions of coastal communities. Furthermore, the AI-based technology in the assessment allows for adjustments to questions or scenarios based on individual student needs, understanding levels, and cultural background. AABAI will also automatically analyze student assessment results and quickly provide in-depth and accurate feedback. This will increase the effectiveness and efficiency of teachers in conducting assessments. In addition, AABAI also utilizes multimedia elements (video, images, audio) to be more contextual in describing the culture of coastal communities, coastal environments, or other coastal activities associated with science materials. Automatically, AABAI can develop critical thinking skills and solve existing environmental problems.

This research aims to develop and validate an AI-based automatic assessment instrument (AABAI) to measure the environmental literacy of junior high school students, particularly in coastal settings. The study seeks to (1) construct a valid and reliable AABAI instrument that aligns with ecological and cultural contexts; (2) assess its content validity, empirical validity, and reliability; and (3) explore profile of students' critical thinking.

2. Method

This research employs the ADDIE framework within a Research and Development (R&D) approach. The primary objective of this study is to develop a product in the form of an assessment instrument designed to measure students' environmental literacy. The research method in this study is an adoption of a combination of the ADDIE model Branch (2019) and development model Oriondo & Dallo-Antonio (1984) includes test planning, trials, validation, reliability assessment, and interpretation of test scores. (Salleh & Ong, 2019).

The main product of this study is AABAI, which can be used to diagnose students' environmental literacy in junior high school science subjects, namely topics related to ecological issues. The data were analyzed using a combination of descriptive quantitative and qualitative methods. The qualitative analysis focused on evaluating the instrument's design based on expert judgment to ensure its conceptual validity. Quantitative analysis is used to determine the validity and reliability of the instrument. The instrument design is made and developed into questions, then given to experts for content and media validity. The test instrument consists of 20 with descriptions that vary in the form of complex multiple-choice questions, short descriptions, long descriptions, case studies, and project assignments. Each environmental literacy indicator is adjusted to the cognitive level at levels C4 (analysis), C5 (evaluation), and C6 (creating). This cognitive level is the level of mental development of grade 9 students.

The test questions also integrate the local culture of coastal communities as a focus for solving problems in the coastal environment. Coastal environmental issues were chosen because they are essential nationally and globally. Coastal areas are strategic zones that support biodiversity, food security, and the lives of millions of people but are highly vulnerable to the impacts of climate change and unsustainable resource exploitation. Students' environmental literacy indicators were adopted based on Santosa et al (2021). All tests created have been able to measure the four selected environmental literacy indicators.

a. Participant

Participants in this study consisted of 242 ninth-grade students from three public junior high schools in Bengkulu, Indonesia, during the 2024/2025 academic year. The selected schools were: SMP Negeri 19 (three classes), SMP Negeri 21 (three classes), and SMP Negeri 3 (one class). These schools were intentionally selected using purposive sampling to represent a range of school performance levels categorized as low, medium, and high based on the schools' average scores in summative science assessments over the past two years (2022–2024).

In addition to academic level, the sample considered diversity in gender, socio-economic background, and students' prior environmental understanding, which was measured using a short diagnostic questionnaire administered before the main instrument. The average age of participants was 14 years, with a relatively balanced distribution between male and female students.

The empirical validation of the AABAI instrument was conducted through a one-time testing design (cross-sectional) without pre-test or post-test. The focus of this validation phase was to examine the instrument's item validity, reliability, and response patterns using Rasch modeling, not to measure learning gains over time. The choice of coastal schools was intentional, given that Bengkulu is located along the western coastline of Sumatra, where coastal community culture remains strong. Embedding coastal environmental issues in the instrument aims to create a more contextual and meaningful learning experience for students.

b. Instrument

The instrument, designed to serve as an indicator of environmental literacy, underwent content validation through expert evaluation. A panel of six assessors participated in the validation process, comprising two specialists in educational assessment, two experts in science education, and two field practitioners. In addition, three media experts evaluated the AABAI instrument. To assess its readability, the instrument was administered to a sample of 242 students. Subsequently, item trials were conducted to gather empirical evidence supporting the instrument's quality.

c. Procedures

The complete AABAI development procedure is shown in Figure 1.

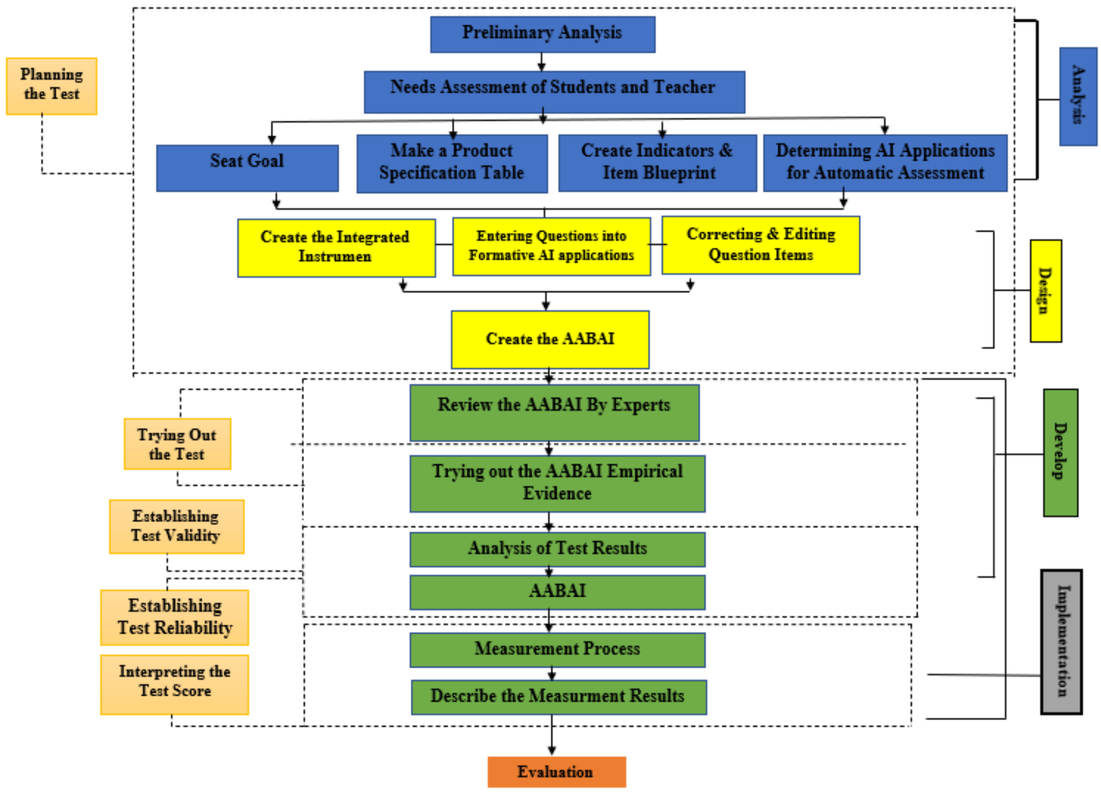

Figure 1. AABAI Development Procedure

The formative application used in this study is an AI-based automatic assessment platform called Formative. This platform was selected for its real-time feedback capabilities, the ability to monitor student learning dynamically, and its varied question formats, including multimedia integration. These features make it particularly suitable for embedding contextualized science problems related to coastal environmental issues. The menus available for displaying questions are also very varied. It can display images, videos, graphics, sound, and simulations so that students avoid answering questions carelessly. Question types also vary: multiple choice, complex multiple-choice, case studies, short descriptions, long descriptions, answering with audio, uploading images, and many more. This variation in the appearance of questions will minimize students answering carelessly.

However, the implementation of AI-based tools in education raises concerns regarding algorithmic bias, particularly when applied to culturally diverse student populations (Blodgett et al., 2020; Ferrara, 2024). To address this, several mitigation strategies were applied. First, the content used in the assessment was localized to reflect the cultural and environmental context of coastal communities in Bengkulu, including traditional marine practices and local environmental challenges. This adaptation ensures that language, examples, and values embedded in the tasks are relevant and inclusive.

Second, while the AI features of Formative facilitate automatic scoring and feedback, teachers retained control over final interpretations, especially in responses that may involve cultural reasoning, local dialects, or non-standard expressions that AI might misinterpret. This hybrid model combining AI automation with teacher judment reduces the risk of misclassification and cultural insensitivity (Williamson & Eynon, 2020).

This approach aligns with the principles of AI fairness in education, which emphasize the need for transparency, contextual relevance, and human oversight to prevent reinforcing systemic inequalities (Luckin et al., 2019). By integrating local context and maintaining teacher involvement, the use of Formative in this study supports equitable and culturally responsive assessment practices.

After the instrument is ready, the validity and reliability of the test questions are tested. The content validity test is given to material and instrument experts to assess suitability, completeness, and readability. Media validity is given to media experts to determine the quality of AI and functionality, display design, usefulness in learning.

d. Data Analysis

The V Aiken formula determines the validity of the test instrument's content. The content validity coefficient is based on expert assessments. The V Aiken index value is calculated as follows:

$$V = \frac{\sum s}{[n(c-1)]} = \frac{\sum(r - lo)}{[n(c-1)]}$$

Where s is the assessment score minus the lowest score on the scale, n is the number of experts, and c is the number of categories on the assessment scale. The V value is between 0 and 1, where the closer it is to 1, the higher the level of content validity. (Aiken, 1985; Azwar, 2012). According to Azwar (2015), the validation criteria for the AABAI instrument in assessing environmental literacy are classified into five distinct categories, as shown in Table 1.

Table 1. Criteria Value of Aiken's Validity

Validity Value	Category
$0.8 \leq V \leq 1.0$	Very Good
$0.6 \leq V \leq 1.8$	Good
$0.4 \leq V \leq 0.6$	Acceptable
$0.2 \leq V \leq 0.4$	Bad
$V \leq 0.2$	Very Bad

The validation of the AABAI instrument was further supported by empirical data obtained through the analysis of test item responses using polytomous data. This data was examined using Item Response Theory (IRT), specifically employing the Rasch model or the Partial Credit Model (PCM) based on the one-parameter logistic (1-PL) approach. The analysis was conducted with the assistance of Quest and Microsoft Excel software. Quest was utilized to assess item fit statistics, test reliability, and item difficulty indices. The Excel program was used to display information on student ability profiles. According to Hambleton and Swaminathan (1985), test items are considered to align well with the Partial Credit Model (PCM) when the INFIT Mean Square (MNSQ) values fall within the range of 0.5 to 1.5, and the INFIT t-statistics lie between -2.0 and 2.0. The reliability of the integrated assessment instrument can be calculated using the Quest software. The reliability index is obtained from the output file with the .sh extension, specifically found in the Item Estimation Summary section. The higher the reliability coefficient, the more reliable the instrument and the smaller the possibility of error (Subali, 2015). George & Mallery

(2020) categorize the reliability coefficients as shown in Table 2.

Table 2. Interpretation of Reliability

Reliability Coeficient	Category
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$V < 0.5$	Unacceptable

Student feedback on the readability of the instrument, as product users, was collected through a questionnaire that had been previously validated by expert reviewers.

The resulting scores were then translated into feasibility categories based on the criteria outlined in Table 3. (Sumadi et al., 2015).

Table 3. Criteria Value of Readability

No	Score range	Category
1	$Xi + 1.8\ Sbi < X$	Very Good
2	$Xi + 0.6\ Sbi < X \leq Xi + 1.8\ Sbi$	Good
3	$Xi - 0.6\ Sbi < X \leq Xi + 0.6\ Sbi$	Acceptable
4	$Xi - 1.8\ Sbi < X \leq Xi - 0.6\ Sbi$	Bad
5	$X \leq Xi - 1.8\ Sbi$	Very Bad

The level of difficulty of the environmental literacy test items was determined by examining the estimate (b) values, as presented in Table 4 (Bond & Fox, 2015).

Table 4. Difficulty Level

Difficulty Level	
$b > 2$	Very Dificult
$1 < b \leq 2$	Dificult
$-1 \leq b \leq 1$	Moderate
$-1 > b \geq -2$	Easy
$B < -2$	Very Easy

Meanwhile, the media validity coefficient is based on expert assessments. While the readability of the text of the reference score change instrument is in Table 5 (Mardapi, 2008).

Table 5. Readability Score

No	Interval	Score Range	Category
1	$X \geq \bar{\bar{X}} + 1.SB_X$	$X \geq 3.0$	Very Good
2	$\bar{\bar{X}} + 1.SB_X > X \geq \bar{X}$	$2.5 \leq X < 3.0$	Good
3	$\bar{\bar{X}} > X \geq \bar{X} - 1.SB_X$	$2.0 < X < 2.5$	Acceptable
4	$X < \bar{\bar{X}} - 1.SB_X$	$X \leq 2.0$	Very Bad

3. Result and Discussion

a. Content Validity

The content validity of the test instrument developed in this study was evaluated by six expert assessors using a four-point rating scale. According to Aiken's criteria, the minimum acceptable value for the Aiken's V coefficient in this context was 0.78 at a 0.05 level of significance (Aiken, 1985). The results of this validation analysis are illustrated in Figure 2.

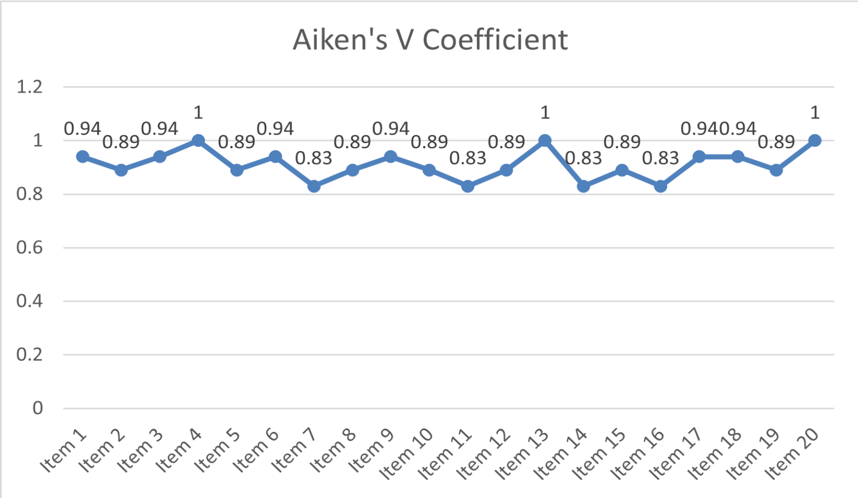

Figure 2. Student Environmental Literacy Scores Based on Aiken's V Coefficient

b. Empirical Validity

The analysis results show that the estimated item reliability is in the range of 0.81 to 1.23, which means that the sample is consistent with the items being evaluated and is very good, as seen in Table 6. The results of the goodness-of-fit analysis, based on the INFIT Mean Square (MNSQ) parameters, demonstrate that the AABAI instrument used to assess environmental literacy satisfies the statistical fit requirements, as detailed in Table 6.

Table 6. Fit Model at 0.5 Probability Level

No item	INFIT MNSQ	Criteria
1	1.00	FIT
2	1.01	FIT
3	0.99	FIT
4	1.02	FIT
5	0.95	FIT
6	0.96	FIT
7	1.02	FIT
8	0.98	FIT
9	0.99	FIT
10	0.98	FIT
11	1,03	FIT
12	1.00	FIT
13	1.01	FIT
14	1.02	FIT
15	1.01	FIT
16	0.97	FIT
17	0.99	FIT
18	1.05	FIT
19	1.04	FIT
20	0.98	FIT

Figure 3 presents a set of 20 items analyzed using the Rasch model. According to Hambleton and Swaminathan (1985), an item is considered to have acceptable quality if its difficulty index falls within the range of -2 to +2 logits, indicating an appropriate level of difficulty.

The analysis presented in Figure 3 indicates that all 20 items are classified within the acceptable or "good" category based on their difficulty levels.

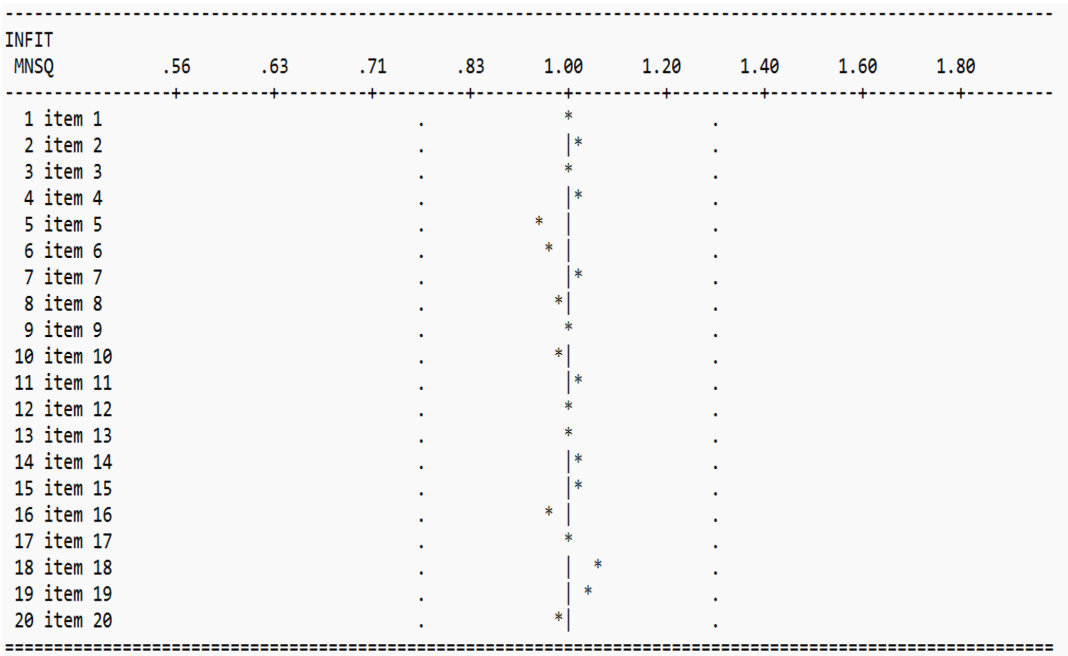

Figure 3. Map fit model

c. Level of Difficulty

An effective test item is one that satisfies the criteria for both validity and reliability, and possesses an appropriate level of difficulty. Table 7 presents the results of the analysis conducted to determine the difficulty levels of the test items.

Table 7. Level of Difficulty Environmental Literacy Questions

No. Item	Difficulty (b)	Difficulty Category
1	-0.23	Moderate
2	-1.24	Easy
3	1.71	Difficult
4	1.53	Difficult
5	-0.28	Moderate
6	-0.22	Moderate
7	-0.16	Moderate
8	-0.08	Moderate
9	-1.75	Easy
10	-0.14	Moderate
11	-0.17	Moderate
12	-0.20	Moderate
13	-0.18	Moderate
14	-0.17	Moderate
15	-0.11	Moderate
16	-0.23	Moderate
17	-0.29	Moderate
18	-0.28	Moderate
19	-0.15	Moderate
20	-0.10	Moderate

Table 7 shows that 16 questions (80%) fall into the moderate category, 2 questions (10%) fall into the difficult category, and 2 (10%) questions fall into the easy category.

d. Media Validation Test

The results of the media validation test (AI quality and functionality, display design (user interface), usefulness in learning) involving three media validators. Figure 4 shows the results of the media validation on the criteria of "very good" without revision. Therefore, this test instrument is suitable for use.

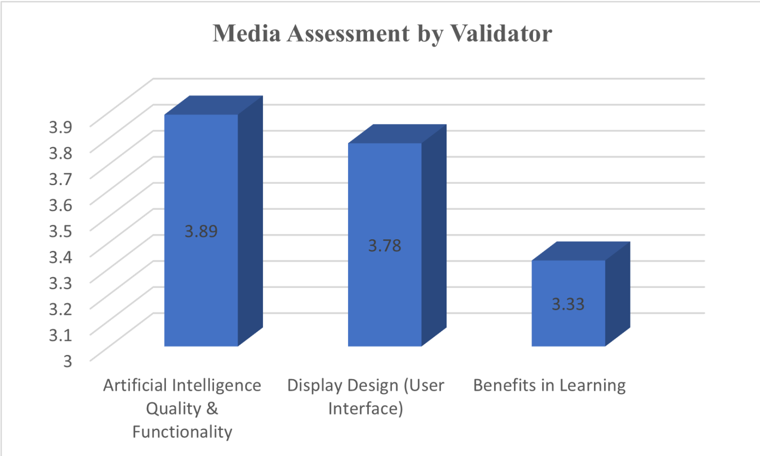

Figure 4. Media Assessment by Validator

e. Readability of Question Text

The purpose of the readability test is to assess the complexity level of the written text. This evaluation provides an estimate of readers' comprehension at the sentence level. Readability, which reflects how easily or difficultly the content of a text can be understood, is illustrated in Figure 5.

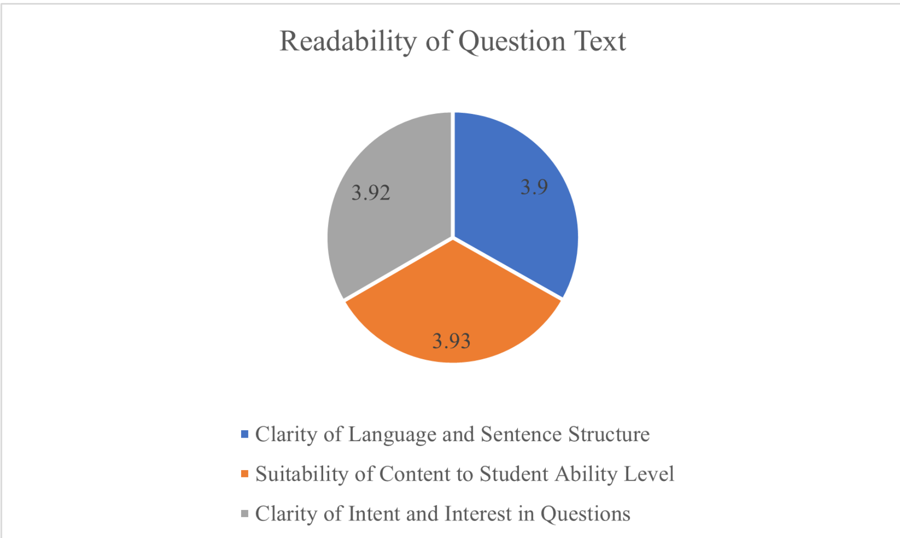

Figure 5. Readability of Question Text

The results of the instrument readability test (seen from the aspects of clarity of language and sentence structure, suitability of content to student ability level, and clarity of intent and interest in questions) involving 242 students showed "very good" results with a slight revision of the simplification of terms. After being revised, this test instrument is suitable for use.

f. Student environmental literacy profile

Based on the environmental literacy test results, most students are in the moderate category (95%), showing a good basic understanding of environmental issues such as waste and global warming. Meanwhile, 5% showed high environmental literacy scores.

The AABAI empirical test was conducted to determine the suitability of environmental literacy indicators with the formulated questions. This test instrument focuses on environmental issues as the primary material, which is then divided into sub-topics such as food availability, environmental health, energy crisis, and global warming. This topic is still limited but promising, especially in the context of strengthening students' awareness and competence towards environmental issues (Syahmani et al., 2021). This study uses environmental literacy indicators based on Santosa et al (2021), then integrated with local cultural traditions of coastal communities so that environmental issues are more contextual with a focus on coastal areas and adapted to the cognitive abilities of C4-C6.

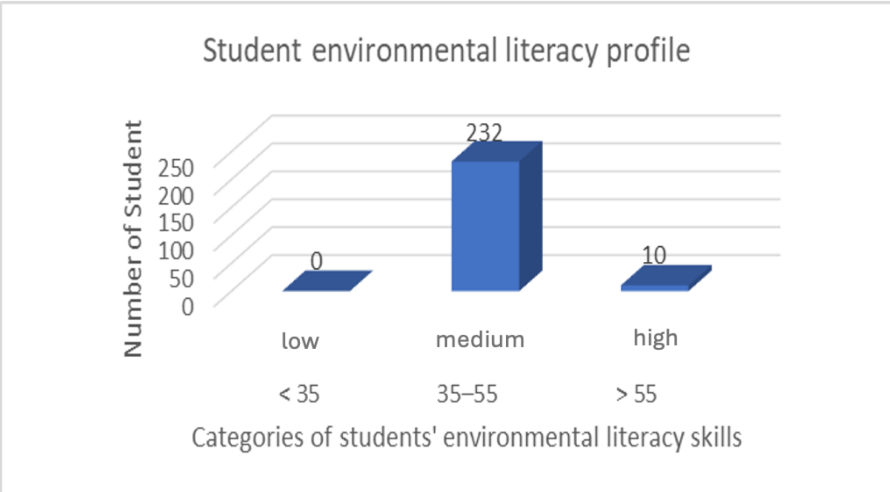

Figure 6. Student Environmental Literacy Profile

The content validity of the developed test instrument was reviewed by six expert validators using a four-point assessment scale. This validation step was conducted prior to empirical testing. Following Aiken's (1985) criteria, a minimum acceptable Aiken's V coefficient of 0.78 at the 0.05 significance level was adopted. As presented in Figure 2, all item scores ranged from 0.98 to 1.05, indicating that each item significantly exceeded the threshold and thus fulfilled the content validity criteria. Based on the classification in Table 1, all items fall into the "very good" category.

These findings are consistent with previous studies, such as García-Ceberino et al (2020), which stated that items with an Aiken's V value ≥ 0.77 are considered valid, and Mistiani et al (2022), who emphasized that instruments with V values above 0.80 indicate high content validity.

However, such uniformly high values across all items raise important questions about the underlying context. Several factors may have contributed to this outcome. First, the items were developed based on science concepts aligned with the existing junior high school science curriculum, especially on topics like waste, pollution, and coastal ecosystems issues that are not only taught in schools but also highly visible in students' daily lives in Bengkulu's coastal environment. This contextual familiarity may have led to clearer item wording and stronger agreement among validators regarding content relevance.

Second, during the development process, items were designed in collaboration with local educators and environmental experts, incorporating authentic and culturally relevant environmental issues (e.g., marine debris, coral reef damage, and traditional coastal practices). This alignment between local context and scientific content likely made the items easier to evaluate positively, both in terms of curriculum relevance and construct representation.

Third, the validators were selected based on expertise in science education, assessment, and environmental literacy, which may have introduced consistency in judgment criteria, leading to less variation in scoring.

While these results are encouraging, they may also suggest that future iterations of the instrument should include more complex or unfamiliar scenarios to assess deeper conceptual understanding or critical thinking, particularly for higher-order cognitive demands. Including items that go beyond recall and recognition could provide a more nuanced picture of environmental literacy levels and challenge students to engage in reflective decision-making.

The high content validity (Aiken's V values ranging from 0.98 to 1.05) and strong empirical reliability of the AABAI instrument are not only statistically significant but also carry important practical implications for implementation in real educational settings. For teachers, the validated and reliable nature of the instrument enhances confidence in using AABAI as a dependable tool for assessing students' environmental literacy. When teachers are assured that each item has been reviewed for clarity, relevance, and representativeness, they are more likely to integrate the assessment into classroom practice without concerns about ambiguity or misalignment with curriculum goals.

Moreover, the reliability indicators both item and person reliability demonstrate that the instrument performs consistently across diverse student groups. This consistency is crucial when used in schools of varying academic levels (low–medium–high) as it ensures measurement fairness and comparability of results across contexts.

Importantly, the instrument was deliberately developed to reflect local environmental and cultural contexts, especially those specific to coastal communities in Bengkulu. Items included themes such as marine debris, traditional fishing practices, coral reef degradation, and community-based coastal conservation rituals. This contextual integration helps ensure that AABAI is not only scientifically accurate but also culturally resonant, allowing it to capture local environmental wisdom that may not be visible in standardized national assessments.

As a result, the instrument's high validity and reliability enable it to function not just as an academic assessment tool, but also as a medium for cultural affirmation helping students recognize the value of their community's environmental knowledge and practices. This could also serve as a stimulus for place-based learning, where science education becomes more meaningful because it is rooted in students' lived experiences.

The results of the test of the level of difficulty of the test questions are shown in Table 5. A total of 16 questions (80%) are in the moderate category, two questions (10%) are in the difficult category, and 2 (10%) questions are in the easy category by looking at the estimated value (b). The distribution of the questions' difficulty level has met the provisions for the distribution of the difficulty level. Data collection was carried out by providing 20 AABAI questions in 60 minutes. When creating the questions, the questions were classified into certain difficulty levels (easy, moderate, and difficult) based on the expected answer requests in AABAI. Questions in the form of projects, which require answers to design/draw/plan a project, are considered the most difficult. Meanwhile, questions with answer requests, such as complex multiple-choice, which only choose more than one correct answer, are considered the easiest questions. A test item is considered to be of good quality when its level of difficulty is balanced neither excessively difficult nor overly simple meaning it falls within the moderate or acceptable range. A test is considered good if it is reliable. As presented in Table 2, the test reliability falls within the "good" category, indicating that the instrument is consistently dependable for assessing students' environmental literacy.

Figure 4 shows that in terms of AI quality and functionality, an average of 3.89 is obtained, which is included in the "very good" category. Regarding display design (user interface), an average of 3.78 is obtained, which is included in the "very good" category. Regarding its effectiveness in the learning process, the AABAI instrument achieved an average score of 3.33, which falls within the "very good" category. This outcome suggests that AABAI fulfills the criteria of a high-quality assessment tool, particularly in terms of artificial intelligence performance, system functionality, user interface accessibility, and its overall contribution to enhancing learning experiences. This indicates that AABAI is worthy of being implemented in the educational context as an assessment innovation that is efficient, adaptive, and responsive to the needs of students. Empirical support from international literature further strengthens these findings. Cui et al (2018) intelligent interface design in an AI-based assessment system increased student engagement by 25% in a field study involving over 20.000 participants. A study of the adaptive learning system "Yixue" in China reported significant improvements in academic performance compared to traditional methods. Awang et al (2024) AI-based adaptive learning platforms increase user motivation, conceptual understanding, and satisfaction. In addition, adaptive AI systems can personalize the learning experience and improve student engagement and performance (Gligorea et al., 2023). The combination of these results shows that AABAI is technically superior and has been proven to contribute positively to the effectiveness of learning and student engagement.

Figure 5 shows that the clarity of language and sentence structure, suitability of content to student ability level, and intent and interest in questions are in the "very good" category. This means that the AABAI assessment instrument has a "very good" level of readability in terms of clarity of language and sentence structure, suitability of content to student ability level, and clarity of purpose and interest in questions. So AABAI is considered effective in conveying the purpose of the evaluation communicatively and easily understood by students and has the potential to increase their involvement in the assessment process. Yaacoub & Prevost (2025) highlighted that language-based AI feedback mechanisms significantly improved the readability and adaptability of questions, resulting in more meaningful student responses. Furthermore, Ali et al (2025) was found that using AI-based assessment tools designed with a focus on readability and language structure effectively strengthened students' positive perceptions of the instruments' validity and relevance.

Figure 6 shows that the majority of students (95%) are categorized as having moderate environmental literacy, indicating that they possess a basic understanding of environmental issues such as waste, global warming, and their ecological and social impacts. This suggests that students are generally aware of key environmental challenges, but have not yet developed the higher-level skills necessary for critical evaluation or transformative action.

One likely reason is the limited contextualization of previous science learning approaches. Based on curriculum observations and teacher interviews, environmental topics such as pollution, climate change, and coastal degradation were often delivered through lecture-based methods, lacking direct connection to students' local environments or cultural experiences. This may have hindered students from transferring conceptual understanding into real-world applications, which is essential for reaching the high literacy level.

Furthermore, students limited exposure to higher-order assessment formats, such as project-based or problem-based learning that require critical thinking, collaborative inquiry, and the ability to connect scientific knowledge with sustainable action. Since most classroom assessments focused on recall and factual understanding, students were less prepared to handle the complex, scenario-based questions embedded in the AABAI instrument.

Only 5% of students fall into the high category, which reflects their ability to not only comprehend environmental problems but also to apply that knowledge in analyzing real-world issues and proposing sustainable solutions. This profile aligns with findings by Örs (2022), who reported that even students with strong environmental attitudes may lack depth in conceptual understanding. Similar trends were reported in the United States by the NOAA's National Environmental Literacy Assessment (NELA), which revealed that most students scored in the moderate range and emphasized the need for context-rich and cognitively challenging instruction (McBeth et al., 2011).

To explore possible factors influencing these results, correlation analysis was conducted between item difficulty levels (as measured using Rasch logit scores) and student performance levels. The results show a positive correlation ($r = 0.41$) between item difficulty and student achievement category, suggesting that students in the high category were more likely to succeed on high-difficulty (higher-logit) items, while those in the moderate category performed well only on low- to medium-difficulty items. This indicates that students' conceptual depth is still limited, and higher-order items remain challenging for the majority.

Furthermore, an analysis across the three sampled schools (classified as low, medium, and high based on academic performance) reveals differences in literacy profiles. In the high-performing school, 12% of students achieved high literacy, compared to just 3% in the medium-performing school, and none in the low-performing school. This suggests that school context and academic culture may play a role in shaping environmental literacy outcomes. Students in more academically advanced schools are likely more exposed to enriched science learning environments, better instructional strategies, and more frequent engagement with environmental themes in both formal and informal learning contexts.

These findings imply that environmental literacy, while generally present at a basic level among students, remains unevenly developed and sensitive to contextual factors such as school quality and instructional approach. To move students beyond the moderate level, there is a need for differentiated, context-specific, and higher-order learning strategies that engage students in problem-solving, action-oriented projects, and critical reflection on real environmental issues in their communities.

While the technical success of AABAI, demonstrated through high validity, reliability, and user adaptability is a crucial foundation, its transformative value lies beyond technical functionality. AABAI is not only designed to measure environmental literacy, but also to stimulate reflection, build awareness, and encourage pro-environmental behaviors among students, especially those in coastal communities vulnerable to ecological degradation.

Through its scenario-based questions, AABAI presents students with authentic environmental challenges that are closely tied to their daily lives, such as plastic pollution on beaches, coral reef damage, and mangrove conservation. These scenarios require students not only to recall facts, but also to reflect on the causes, consequences, and solutions a process that activates higher-order thinking and personal responsibility. This is aligned with the constructivist principle that meaningful learning occurs when students can relate new information to their lived experiences (Piaget, 1972; Vygotsky, 1978).

Moreover, the immediate and personalized feedback generated by AABAI helps students understand the implications of their choices, correct misconceptions, and explore sustainable alternatives. For example, if a student selects an ineffective or unsustainable solution in a coastal scenario, the system explains why it is problematic and suggests community-based alternatives thus reinforcing both cognitive accuracy and behavioral guidance.

Therefore, AABAI plays a dual role: as an assessment tool that measures environmental literacy dimensions (knowledge, attitude, behavior, reflection) and as a learning tool that fosters environmental awareness and motivates students to act. Future studies should include longitudinal designs to better capture AABAI's influence on long-term behavior change and its potential as a catalyst for community-based environmental education.

The novelty of this research lies in the development of AABAI an AI-based assessment tool that uniquely integrates adaptive feedback, environmental literacy dimensions, and local coastal cultural context. Unlike previous studies such as Cui et al. (2018), which focused on general AI-assisted formative assessments, or Awang et al. (2022) and Gligorea et al. (2024), which emphasized standardized instruments with limited cultural depth, AABAI embeds authentic, place-based scenarios rooted in the lived experiences of Bengkulu's coastal communities. These include traditional ecological practices, local environmental challenges, and cultural values related to sustainability. Furthermore, AABAI not only automates scoring but also provides meaningful, real-time feedback tailored to students' responses, encouraging critical reflection and pro-environmental behavior. This combination of automated assessment, cultural contextualization, and pedagogical impact represents a significant advancement over existing instruments, positioning AABAI as a transformative tool that not only measures but also fosters environmental awareness and local action an innovation not yet seen in prior AI-driven environmental literacy research.

4. Conclusion

The results of this study confirm that AABAI is a valid and reliable instrument for assessing junior high school students' environmental literacy in a contextualized coastal setting. Content validation through expert judgment using Aiken's V produced values ranging from 0.98 to 1.05, exceeding the standard threshold of 0.78 and indicating that all items were highly relevant and clearly formulated. Empirical analysis further demonstrated strong internal consistency, with item reliability and person reliability meeting acceptable psychometric standards. In terms of readability, student feedback indicated that the questions were easy to understand and engaging, particularly due to the integration of multimedia elements that reflect local environmental and cultural contexts. The item difficulty distribution also showed a balanced range across cognitive levels, although most students performed better on low-to-medium difficulty items, suggesting the need to strengthen higher-order thinking components.

AABAI contributes significantly to environmental education by not only providing an automated, adaptive, and efficient assessment model but also by embedding cultural relevance and real-world environmental issues that are familiar to students in coastal areas. This dual function measuring and promoting environmental awareness positions AABAI as a transformative educational tool aligned with the goals of contextual learning and sustainability education. The findings also highlight the importance of shifting instructional practices toward more experiential, project-based, and culturally embedded strategies to elevate students from moderate to high levels of environmental literacy.

Future research should explore the long-term impact of AABAI on behavioral change and environmental decision-making among students. Additionally, expanding the instrument to other ecological and cultural contexts beyond coastal communities can strengthen its generalizability. Further development of AABAI could also include adaptive item pathways based on real-time student responses and the integration of student-generated content to enhance engagement and ownership of learning.

5. Funding

The authors is grateful to LPPM Universitas Bengkulu for providing research funds from PNBP Universitas Bengkulu for the Unggulan research scheme with contract number: 1815/UN30.15/PT/2025.

6. References

Afriansyah, A., & Sukmayadi, T. (2022). The local wisdom value of the sea alms tradition in increasing the spirit of mutual cooperation among the coastal communities of Pelabuhan Ratu Beach. *Jurnal Penelitian Ilmu-Ilmu Sosial*, *3*(1), 38-54.

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

Akbar, M., & Maghfira, A. (2023). The Impact of Plastic Waste on Seawater Pollution in Makassar City. *Riset Sains Dan Teknologi Kelautan*, *6*(1), 25–29. https://doi.org/10.62012/sensistek.v6i1.24234

Ali, Z., Bhar, S. K., Abd Majid, S. N., & Masturi, S. Z. (2025). Exploring Student Beliefs: Does Interaction with AI Language Tools Correlate with Perceived English Learning Improvements? *Education Sciences*, *15*(5), 1–18. https://doi.org/10.3390/educsci15050522

Aliviyanti, D., Kasitowati, R. D., Yona, D., Semedi, B., Rudianto, R., Asadi, M. A., & Dewi, C. S. U. (2022). Education on the Dangers of Plastic Waste in Waters and Marine Biota at the Nature School, Bajulmati Beach, Malang Regency, East Java. *Abdi Geomedisains*, 119-129.

Ariyatun, A., Sudarmin, S., Wardani, S., Saptono, S., & Winarto, W. (2024). Bibliometric Analysis of Environmental Literacy in Sustainable Development: A Comprehensive Review Based on Scopus Data From 2013 to 2023. *International Journal of Educational Methodology*, *10*(1), 179–195. https://doi.org/10.12973/ijem.10.1.979

Aryanti, F., Surtikanti, H., & Riandi, R. (2017). Implementation of Problem Based Learning (PBL) assisted by Information and Communication Technology to Improve Students' Critical Thinking Skills on the Concept of Environmental Pollution. *BIOSFER : Jurnal Biologi dan Pendidikan Biologi*, *2*(1), 14–20. https://doi.org/10.23969/biosfer.v2i1.370

Awang, L. A., Yusop, F. D., & Danaee, M. (2024). Insights on usability testing: The effectiveness of an adaptive e-learning system for secondary school mathematics. *International Electronic Journal of Mathematics Education*, *19*(3). https://doi.org/10.29333/iejme/14621

Azwar, S. (2015). *Reliability and Validity (fourth)*. Pustaka Pelajar.

Blodgett, S. L., Barocas, S., Daumé, H., & Wallach, H. (2020). Language (Technology) is Power: A Critical Survey of "Bias" in NLP. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 5454–5476. https://doi.org/10.18653/v1/2020.acl-main.485

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Branch, Robert Maribe. (2009). *Instructional Design: The ADDIE Approach*. New York: Springer.

Cui, W., Xue, Z., & Thai, K. P. (2018). Performance Comparison of an AI-Based Adaptive Learning System in China. *Proceedings 2018 Chinese Automation Congress, CAC 2018*, 3170–3175. https://doi.org/10.1109/CAC.2018.8623327

Dwiyanti Suryono, D. (2019). Plastic Waste in Coastal and Marine Waters: Implications for the Coastal Ecosystem of DKI Jakarta. *Jurnal Riset Jakarta*, *2*(1), 17–23. https://doi.org/10.37439/jurnaldrd.v12i1.2

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1). https://doi.org/10.3390/sci6010003

García-Ceberino, J. M., Antúnez, A., Ibáñez, S. J., & Feu, S. (2020). Design and validation of the instrument for the measurement of learning and performance in football. *International Journal of Environmental Research* *and Public Health*, *17*(13), 1–22. https://doi.org/10.3390/ijerph17134629

George, D., & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference (Sixteenth)."* Routledge. https://doi.org/10.4324/9780429056765

Gligorea, I., Cioca, M., Oancea, R., Gorski, A. T., Gorski, H., & Tudorache, P. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. *Education Sciences*, *13*(12). https://doi.org/10.3390/educsci13121216

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences (Switzerland)*, *11*(12). https://doi.org/10.3390/app11125467

Hambleton, R. K., & Swaminathan, H. (1985). *Items Response Theory: Principles and Application.* Kluwer-Nijjhoff Publish.

Hamid, T., Chhabra, M., Ravulakollu, K., Singh, P., Dalal, S., & Dewan, R. (2022). A Review on Artificial Intelligence in Orthopaedics. *Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development, INDIACom 2022*, 365–369. https://doi.org/10.23919/INDIACom54597.2022.9763178

Handiyati, T., Qomariyah, S., & Kurniawan, J. (2023). The Role of Environmentally Based Learning in Improving Student Understanding at MI Cimahi Peuntas, Sukabumi Regency. *Jurnal Pendidikan Karakter*, *1*(4), 108–140. https://doi.org/10.2307/j.ctv1nth4c.13

Hohn, S; Acevedos-Trejos, E; Abrams, J. et al. (2013). The long-term legacy of plastic mass production. *Journal of Cleaner Production*, 0–48.

Hungerford, H. R., & Peyton, R. Ben. (1976). A Paradigm of Environmental Action. *Department Health Education*.

Husamah, H., Rahardjanto, A., Hadi, S., Lestari, N., & Ummah BK, M. K. (2023). Spirituality-based environmental literacy among prospective biology teacher in Indonesia: Analysis based on gender, accreditation, and semester-level aspects. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *9*(3), 418–432. https://doi.org/10.22219/jpbi.v9i3.29500

Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., & Law, K. L. (2015). Plastic waste inputs from land into the ocean. *Science*, *7*(6), 768–771. https://doi.org/10.1126/science.1260352

Kahn, R. (2008). From Education for Sustainable Development to Ecopedagogy: Sustaining Capitalism or Sustaining Life? *Green Theory & Praxis: The Journal of Ecopedagogy*, *4*(1), 1–14. https://doi.org/10.3903/gtp.2008.1.2

Krstić, L., Aleksić, V., & Krstić, M. (2022). Artificial Intelligence in Education: A Review. *Technics and Informatics in Education – TIE 2022*, *September*, 223–228. https://doi.org/10.46793/tie22.223k

Law, K. L., Starr, N., Siegler, T. R., Jambeck, J. R., Mallos, N. J., & Leonard, G. H. (2020). The United States' contribution of plastic waste to land and ocean. *Science Advances*, *6*(44), 1–7. https://doi.org/10.1126/sciadv.abd0288

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2019). *Intelligence unleashed: an argument for AI in education. Resource document* (Issue October).

Mardapi, D. (2008). *Techniques for Developing Test and Non-Test Instruments*. Yogyakarta: Mitra Cendekia.

Masrobi, Intan Valentina, Widodo Casraji Putra, Umi Hanni, Maulidiyanto, & Gusman, T. A. (2022). Efforts to Increase Public Awareness of the Importance of Cleanliness of Sarwajala Beach. *BAKTIMU: Jurnal Pengabdian Kepada Masyarakat*, *2*(2), 87–94. https://doi.org/10.37874/bm.v2i2.375

McBeth, B., Hungerford, H., Marcinkowski, T., & Volk, T. (2011). National Environmental Literacy Assessment, Phase Two: Measuring the Effectiveness of North American Environmental Education Programs with Respect to the Parameters of Environmental Literacy. *NOAA and NAAEE ….*

Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100040. https://doi.org/10.1016/j.chbah.2023.100040

Mistiani, W., Istiyono, E., & Syamsudin, A. (2022). European Journal of Educational Research. *European Journal of Educational Research*, *11*(1), 69–81. https://pdf.eu-jer.com/EU-JER_9_1_395.pdf

Oriondo, L. L., & Dallo-Antonio, E. M. (1984). *Evaluating Educational Outcomes (Test, Measurement, and Evaluation) (5th Editio)*. REX Printing Company, Inc, 279-280.

Örs, M. (2022). A Measurement of the Environmental Literacy of Nursing Students for a Sustainable Environment. *Sustainability (Switzerland)*, *14*(17). https://doi.org/10.3390/su141711003

Piaget, J. (1972). *The psychology of the child*. New York: Basic Books.

Puspita, M. (2017). Local Wisdom in Coastal and Marine Resource Management: Customary Law of the Laot and the Panglima Laot Institution in Nanggroe Aceh Darussalam. *Sabda : Jurnal Kajian Kebudayaan*, *3*(2). https://doi.org/10.14710/sabda.v3i2.13253

Putra, H., & Dupuy, K. F. P. (2023). Coastal Zone Management Strategy for Reducing Plastic Waste in Accordance with SDG 14 Marine Ecosystems in the Coastal Area of DKI Jakarta. *Riset Sains dan Teknologi Kelautan*, *6*(2), 178–182. https://doi.org/10.62012/sensistek.v6i2.31701

Salleh, S. M., & Ong, E. T. (2019). The Development and Validation of an Instrument to Measure Higher Order Thinking Skills. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(8 Special Issue), 150–162. https://doi.org/10.36681/tused.2025.004

Samidjo, G. S., Widodo, A. S., Kusumastuti, L., & Suryadin, A. (2023). Environmental Literation through Education at MTs Muhammadiyah Gantung, East Belitung, Bangka Belitung. *Warta LPM*, 184-196.

Santosa, R., Roshayanti, F., & Siswanto, J. (2021). Analysis of Environmental Literacy of Junior High School Students. *JPPS (Jurnal Penelitian Pendidikan Sains)*, *10*(02), 1976–1982.

Subali, B. (2015). Measuring Creativity in Science Process Skills in the Context of Assessment for Learning. *Jurnal Cakrawala Pendidikan*, *1*(1), 130–144. https://doi.org/10.21831/cp.v1i1.4196

Sumadi, C. D., Mulyani, S., & ES, W. A. (2015). Development of Hydrocarbon Compound Game Media in Chemistry Learning at SMA Batik 1 Surakarta and SMA Batik 2 Surakarta. *Jurnal Pendidikan Kimia*, *4*(2), 82–88.

Syahmani, S., Hafizah, E., Sauqina, S., Adnan, M. Bin, & Ibrahim, M. H. (2021). STEAM Approach to Improve

Environmental Education Innovation and Literacy in Waste Management: Bibliometric Research. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, *3*(2), 130–141. https://doi.org/10.23917/ijolae.v3i2.12782

Szczytko, R., Stevenson, K., Peterson, M. N., Nietfeld, J., & Strnad, R. L. (2019). Development and Validation of The Environmental Literacy Instrument For Adolescents. *Environmental Education Research*, *25*(2), 193–210. https://doi.org/10.1080/13504622.2018.1487035

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*(PA), 124167. https://doi.org/10.1016/j.eswa.2024.124167

Widarmanto, N. (2018). Local Wisdom in Fisheries Resource Management. *Sabda : Jurnal Kajian Kebudayaan*, *13*(1), 18–26.

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, *45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Yaacoub, A., & Prevost, L. (2025). Analyzing Feedback Mechanisms in AI-Generated MCQs: Insights into Readability, Lexical Properties, and Levels of Challenge. *Proceedings of the 9th International Conference on Computer, Software and Modeling (ICCSM)*.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*. https://doi.org/10.1155/2021/8812542