

Development of an Environment-Based Assessment Instrument for Scientific Creativity in Science Classes at Junior High Schools

Syinta Nur, *Iwan Wicaksono, and Fadiatul Hasinah Muhammad

Bachelor of Science in Science Education Program, Universitas Jember, Jember,
Indonesia

Center for Foundation Studies, University Teknologi Mara, Selangor, Malaysia

* iwanwicaksono.fkip@unej.ac.id

Article Info	Abstract
Article history: Received: April 20, 2026 Accepted: June 19, 2026 Published: July 28, 2026 Keywords: Environment-based assessment Junior high schools Scientific creativity This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license © 2026 Jurnal ilmiah Pendidikan Fisika	Assessment in science education does not focus solely on cognitive aspects but must also measure students' scientific creativity as it relates to environmental contexts. The purpose of this study is to determine the content validity, item validity, reliability, discriminative power, and difficulty level of the assessment instrument developed. This assessment instrument development study utilized the Borg and Gall development model. This model consists of ten stages; however, this research only applied eight stages, namely identifying potential problems, collecting information, product design, product validation, product revision, small group testing, and field testing. The subjects of this study were 105 seventh-grade students from classes A, B, and C at Junior High School. The results of the validation of the developed product by three expert validators showed an average score of 73.2%, categorized as valid; the item validation test showed a calculated r value > the table r value, thus confirming validity. The reliability of the items was 0.78, categorized as reliable. Analysis of the items' discriminative power showed that 28.57% of the items were in the excellent category, 42.86% were in the good category, and 28.57% were in the fair category. The distribution of item difficulty levels was 42.86% easy, 28.57% moderate, and 28.57% difficult. The result of the study shows that the assessment instrument developed has good characteristics and is suitable for measuring scientific creativity in science education.
To cite this article: Nur, S., Wicaksono, I., & Muhammad, F. H. (2026). Development of an environment-based assessment instrument for scientific creativity in science classes at junior high schools. <i>Jurnal Ilmiah Pendidikan Fisika</i>, 10(2), 263-277.	

INTRODUCTION

In 21st-century education, natural sciences is an educational approach that connects individuals with events and phenomena occurring in the natural world around them. It is not merely about mastering information, concepts, and foundational principles, but rather a systematic process for delving deeper into the universe and discovering new things. Natural Sciences (IPA) plays a vital role in shaping a generation capable of critical and creative (Akbar et al., 2025; Indiana et al., 2024). In the midst of rapid advancements in society and technology, mastering scientific concepts alone is not enough. The ability to apply scientific knowledge to find innovative solutions to existing problems known as scientific creativity has become an essential skill.

According to Wicaksono et al., (2020), students with creative thinking skills are able to generate various ideas, solutions, and new approaches to problem-solving. Based on existing research, assessing students' creative thinking skills in science still faces challenges; one of these is the lack of scientific creativity tests specifically designed for

science curriculum content ([Fadllan et al., 2022](#)), which means that students' potential for scientific creativity cannot yet be measured optimally.

Creativity is a person's ability to create something new; it is a combination of previously acquired data or information that is realized in the form of ideas or tangible works ([Sudarti, 2020](#); [Widyasari, 2023](#)). Creativity in science is also known as scientific creativity. Scientific creativity is an essential component of science education, as it enables students to solve problems, contribute their own ideas, develop scientific concepts, and make informed decisions when facing situations involving science-related matters ([Juliastari et al., 2022](#)).

Scientific creativity is defined as the ability to generate valid and relevant original ideas within a specific scientific discipline, where evaluation takes into account scientific knowledge, skills, and scientific frameworks ([Tiana & Wahyuni, 2025](#); [Xu et al., 2025](#)). [Hu and Adey \(2002\)](#) state that scientific creativity is the ability to generate new ideas that are original, flexible, and useful by utilizing scientific knowledge and skills. There are seven indicators of unusual use, real advance, technical production, scientific imagination, scientific problem-solving, creative experimentation, and scientific product. According to [Wicaksono dan Erlina \(2024\)](#), scientific creativity integrates two elements: creativity and scientific knowledge; therefore, it requires specific measurement tools distinct from those used to assess general creativity. Therefore, specially designed assessment instruments are crucial for objectively and comprehensively measuring scientific creativity. Scientific creativity is an element of higher-order thinking skills that students must possess to effectively address real-world environmental challenges in their surroundings ([Kareem, 2022](#)). The scientific creativity models frequently cited by educational researchers include Hu Adey's Scientific Structural Creativity Model (SSCM) (2002), Son's Scientific Creativity Model (2009), and Park's Model of Scientific Creativity (2010) ([Sidek et al., 2020](#)). The following are seven indicators of scientific creativity according to [Hu and Adey \(2002\)](#) as given in Table 1.

Table 1. Indicators of scientific creativity

Indicators of scientific creativity	Objective
Unusual use	Measuring fluency, flexibility, and originality in the use of objects for scientific purposes
Real advance	Measuring fluency, flexibility, and originality in responses to scientific problems
Technical production	Measuring fluency, flexibility, and originality in technical product development
Science imagination	Measuring fluency, flexibility, and originality in scientific imagination
Science problem-solving	Measuring flexibility and originality in solving scientific problems
Creative experimental	Measuring flexibility and originality in creative experimentation
Science product	Measuring flexibility and originality in the design of science products

An assessment instrument is defined as a tool used to obtain valid and reliable information regarding students' abilities in specific areas, including scientific and creative thinking skills ([Faiz et al., 2021](#); [Hu & Adey, 2002](#)). In the context of science education, assessment instruments must not only measure students' understanding of scientific concepts but also assess their critical thinking skills and creativity. With the right instruments, teachers can assess the extent to which students are able to develop ideas, solve problems, and apply knowledge creatively. According to several experts, there are

four criteria for a good test: validity, reliability, difficulty level, and discriminative power ([Augustia et al., 2025](#); [Moto et al., 2022](#)).

One effective approach to science education is the integration of an environmental approach. This approach integrates scientific knowledge with an understanding of and concern for the environment so that students not only learn about theory but also how to adopt appropriate attitudes and strategies in addressing environmental challenges in their daily lives ([Sonra & Suriani et al., 2025](#)). An environmental approach integrated into learning can serve as a solution to address students' learning difficulties, make learning more engaging, reduce excessive use of technical jargon, enhance focus, and improve student understanding; consequently, learning objectives can be achieved more effectively ([Jannah et al., 2022](#)). One science topic that can be linked to environmental issues is matter and its changes. This topic covers discussions on the states of matter and particle models, changes in the states of matter, physical and chemical changes, and the density of matter. The topic of matter and its changes is closely related to the environment, as the natural phenomena and everyday occurrences we observe are part of both physical and chemical changes; thus, the integration of environmental issues into science education is crucial.

Given the challenges and the importance of assessment instruments in the learning process, the researcher aims to develop an environment-based assessment instrument for scientific creativity in science education. An environment-based assessment instrument was chosen with the goal of addressing the challenges faced by students, as no specific assessment of scientific creativity has been conducted to date. Based on the problem description above, the author conducted a study to test the validity, reliability, discriminant power, and level of difficulty of the assessment instrument that was developed.

METHODS

This study is a type of development research based on the Borg and Gall (1983) development model. The Borg and Gall (1983) development model used in this study consists of eight stages: potential problems (identifying problems through interviews and needs analysis), information gathering (gathering information from the literature and school data relevant to the research), product design (designing an environment-based assessment instrument for scientific creativity), product validation (the product is validated by subject matter, media, and language experts), product revision (refining the product based on the validation results), small group pilot test (testing the product on 30 seventh-grade students), product revision (refining the product based on the results of the small group pilot test), and field test (testing the product in classes VIIA, B, and C at Junior High School). This study was conducted at a junior high school. The technique used to select the research subjects was purposive sampling, in which subjects were chosen based on criteria relevant to the research objectives. One of the criteria for selecting research subjects was that the selected class had completed the material on substances and their changes. The subjects of the study were 105 seventh-grade students from classes A, B, and C during the first semester of the 2025/2026 academic year. This study used a validation sheet as a research instrument and an environment-based test consisting of 7 essay questions covering the main topic of matter and its changes. The data analysis techniques are as follows.

The first data analysis technique is validity analysis, which consists of two types: expert validation and item validation. In expert validation, three validators assess the developed instrument and provide comments and suggestions. After the collected data is analyzed, the results are used as the basis for revising the assessment instrument. The data is then calculated using the following formula:

$$Va = \frac{TS_e}{TS_h} \times 100\% \quad \dots (1)$$

Notes:

Va = Expert validation percentage

TSe = Raw score obtained

TSh = Highest score on the questionnaire

The expert validation percentage obtained can be used to determine the level of validity for each aspect of the criteria being evaluated. Table 2 shows the validity levels.

Table 2 Criteria for interpreting validity

Evaluation (100%)	Category
$25 < Va \leq 40$	Invalid
$41 < Va \leq 55$	Partially valid
$56 < Va \leq 70$	Fairly valid
$71 < Va \leq 85$	Valid
$86 < Va \leq 100$	Highly valid

(Azizah et al., 2023)

The next step is item validity analysis. The validity of each item in the assessment instrument must be tested to ensure that it accurately describes the data from the variables being analyzed. An assessment instrument can be considered valid if it measures what it is intended to measure (Adetya et al., 2025). Item validity can be analyzed using the following formula:

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}} \quad \dots (2)$$

Notes:

r_{xy} : Pearson correlation coefficient results

N : Total number of participants

X : Total score for each item

Y : Sum of total scores for all items

The obtained r_{xy} values will be analyzed using Pearson's correlation coefficient, with r table = $r_{(a,n-2)}$. The criteria for interpreting item validation are listed in Table 3.

Table 3 Criteria for the validation of test items

r_{xy}	Category
$r_{\text{calculated}} \geq r_{\text{table}}$	Valid
$r_{\text{calculated}} \leq r_{\text{table}}$	Invalid

Hamidah & Wulandari (2021)

The third stage is reliability. Reliability refers to the property of an instrument that yields stable (consistent) measurement results (Sugiono et al., 2020). Reliability testing is conducted by assessing internal consistency using Cronbach's alpha, as shown below:

$$ri = \left(\frac{k}{k} - 1\right)(1 - \Sigma \sigma b^2 / \sigma^2) \quad \dots (3)$$

Notes:

Ri : instrument reliability

K : number of items

σb^2 : item variance

σ^2 : total variance

The criteria for reliability assessment result in Table 4.

Table 4 Criteria for the reliability of assessment instruments

Coefficient Interval	Category
< 0.20	Very low reliability
0.20 - 0.39	Low reliability
0.40 - 0.59	Moderate reliable
0.60 - 0.79	High Reliable
0.80 - 1.00	Very high reliable

[Genarsih & Tisngati \(2024\)](#)

The next step is to analyze the item discrimination index. Discrimination refers to the ability of a test item to distinguish between students who have mastered the material and those who have not ([Dahlan et al., 2024](#)). The following formula is used to calculate the item discrimination index:

$$DP = \frac{\text{Mean of the top group} - \text{Mean of the lower group}}{\text{Maximum score for the question}} \quad \dots (4)$$

Table 5 lists the criteria for the discriminant index:

Table 5 Item discrimination criteria

Index	Category
0.40 - above	Excellent
0.30 - 0.39	Good
0.20 - 0.29	Fair
0.19 - below	Poor

[Hamidah & Wulandari \(2021\)](#)

The final step is the analysis of item difficulty. Difficulty is a measure of an item's difficulty level, expressed as a numerical value ([Fitriani, 2021](#)). Difficulty is related to discriminative power ([Nasruddin et al., 2021](#)). The formula for determining item difficulty is:

$$P = \frac{\text{mean}}{\text{maximum score set}} \quad \dots (5)$$

The difficulty scores obtained for each item will be analyzed and categorized based on the scores in Table 6.

Table 6 Criteria for interpreting item difficulty levels

Difficulty Index Value	Category
$0.70 < P \leq 0.90$	Easy
$0.30 < P \leq 0.70$	Moderate
$0.20 < P \leq 0.30$	Difficult

[Judijanto et al., \(2025\)](#)

RESULTS AND DISCUSSION

This study produced an assessment instrument to measure environment-based scientific creativity in one of the science subjects taught in junior high school, namely substances and their changes. The study was conducted at a junior high school during the first semester of the 2025/2026 academic year. The research subjects were classes VII A, B, and C, with a total of 105 students. The development of the scientific creativity assessment instrument followed the Borg and Gall development model, which consists of 8 stages. The following

is a detailed description of the steps involved in developing an instrument for assessing scientific creativity based on the Borg and Gall model:

Potential problems

The initial stage of this development research involved analyzing the issues present at junior high school. This stage began with interviews to gather information regarding the implementation of assessment in science education, particularly as it relates to environment-based scientific creativity. Based on the results of interviews conducted with 7th-grade science teachers at Junior High School, it was found that teachers have attempted to link lesson materials to the students' surrounding environment as a context for learning. The environment holds significant potential for use in science education because it is closely tied to students' daily lives. However, the utilization of the environment has not been accompanied by the use of assessment instruments specifically designed to measure scientific creativity, so this ability has never been measured before. The assessment instruments used by teachers are still dominated by questions focused on concept mastery and cognitive learning outcomes and still refer to questions found in textbooks.

Collecting information

The second stage involves gathering information through literature review and needs analysis. The literature review was conducted to obtain information by examining various relevant sources, such as books, journal articles, and previous research findings on scientific creativity, science learning assessment, and environment-based learning. The results of the literature review revealed information on the concept of scientific creativity according to [Hu & Adey \(2002\)](#), who state that scientific creativity is the ability to generate new ideas that are original, flexible, and useful by applying knowledge. Additionally, information was obtained regarding scientific skills, indicators of scientific creativity, and the characteristics of a good assessment instrument suitable for science learning in junior high school. The interview results indicated that the curriculum used in Grade 7 at junior high school is the Merdeka Curriculum. Subsequently, the researcher observed the teaching modules used by the science teacher. This activity was conducted to gain an understanding of the competencies that Grade 7 students at junior high school are expected to achieve. The results of the analysis of these teaching modules were then used as a reference in the development of the assessment instrument, ensuring that the instrument aligns with the instructional materials used and the students' needs. Based on field observations and interviews with teachers, the researcher selected junior high school as the research site. This selection was based on its suitability for the development of an environment-based scientific creativity assessment instrument.

Product design

The third stage is product design. This stage was carried out using Canva to make the questions more appealing through color schemes, images, and elements suitable for junior high school students. Product development began with creating a question outline containing a table of learning outcomes, scientific creativity indicators, question indicators, assessed aspects, measured objectives, question level formulations, cognitive levels, question formats, and question numbers. The assessment instrument for scientific creativity consists of 7 questions, aligned with the number of scientific creativity indicators as outlined by [Hu & Adey \(2002\)](#). The questions are in essay format. This format was chosen to provide students with the space to generate, develop, and communicate new ideas and innovative solutions within a scientific context. The following is an example of Figure 1 showing the results of the grid design.

KISI-KISI SOAL SCIENTIFIC CREATIVITY

Satuan Pendidikan : SMPN 2 Kalisat
 Mata Pelajaran : Ilmu Pengetahuan
 Alam Kelas/Semester : VII/Ganjil
 Materi Pokok : Zat dan Perubahannya
 Bentuk Soal : Uraian
 Waktu : 1 x 80 menit

Capaian Pembelajaran	Indikator Scientific Creativity	Indikator Soal	Aspek	Tujuan yang diukur	Rumusan Soal	Alternatif Jawaban	Level Kognitif	Nomor Soal
Peserta didik memahami sifat dan karakteristik zat, perubahan fisika dan kimia, serta pemisahan campuran sederhana. Peserta didik menerapkan keterampilan inkuiri untuk menyelidiki fenomena terkait zat	Unusual Use	Siswa mampu menyebutkan berbagai ide terhadap benda padat (botol bekas) berdasarkan sifat zat padat.	1. <i>Fluency</i> 2. <i>Flexibility</i> 3. <i>Originality</i>	Pemanfaatan objek untuk tujuan ilmiah	Perhatikan gambar di bawah ini!  Gambar di samping menggambarkan botol bekas minuman yang banyak ditemukan di lingkungan sekolah dan rumah. Sebagian	- Pot tanaman gantung dari botol bekas (<i>fluency, flexibility</i>) - Penampung air hujan sederhana untuk menyiram tanaman (<i>fluency</i>) - Lampu hias dari botol berisi air dan pewarna (<i>originality, flexibility</i>) - Alat penyiram otomatis dengan melubangi tutup botol dan menanamkannya terbalik di tanah (<i>flexibility, fluency</i>)	C4	1

Figure 1 Question outline

After designing the test blueprint, the questions were imported into Canva for editing to make them more visually appealing. The process of creating the test booklet began with designing a cover that included the title of the material to be assessed and an image relevant to the content. The development of the questions resulted in a paper-based test where answers are written directly on the test booklet. The test booklet contains a cover, identification page, and instructions for completing the questions. An example of the test sheet is shown in Figure 2.



Figure 2 Question sheet

Product validation

The next step is to have the assessment instrument validated by experts before it is pilot-tested on students. Product validation is conducted to test the validity of the assessment instrument and to obtain suggestions for its improvement. The environment-based scientific creativity assessment instrument was evaluated by three validators, namely three science education lecturers from the University of Jember. The level of validity of the assessment instrument was assessed by the validators using a validation sheet that evaluated three aspects: content, content construction, and language. The scores given by the validators were processed using Microsoft Excel and analyzed to determine the level of validity of the environment-based scientific creativity assessment instrument. The results of the expert validation analysis by the three validators are presented in Table 7.

Table 7. Results of the expert validation analysis

No	Assessment Aspect	Validator Percentage (%)			Percentage (%)	Criteria
		1	2	3		
1	Content	80	84.5	25	63.1	Sufficiently Valid
2	Construction	76	92.4	56	74.8	Valid
3	Language	72.8	92.8	80	81.8	Valid
Average		76.2	89.9	53.6	73.2	Valid

The results of the expert validation analysis of the scientific creativity assessment instrument in Table 6 show that the content aspect scored 63.1%, categorized as “moderately valid.” The construct aspect scored 74.8%, categorized as “valid,” and the language aspect scored 81.8%, categorized as “valid.” Additionally, the average percentage from the three validators who assessed the three aspects yielded a result of 73.2% in the “valid” category, indicating that the environment-based scientific creativity assessment instrument can be pilot-tested with minor revisions based on the validators’ suggestion.

Product revision

The assessment instrument, which has been validated by three validators, will be revised to improve the quality of the developed product. The product revision involves refining the scientific creativity assessment instrument in accordance with the suggestions and comments from the validators. The objective of the product revision phase is to produce a valid product suitable for pilot testing on research subjects. Some questions require the addition of an instruction to look at the picture first before answering, some questions contain conflicting instructions, the wording of some questions needs to be revised to eliminate ambiguity and ensure clarity, and several words need to be corrected due to typos. The revised assessment instrument can be considered valid.

Small group testing

The small-group pilot test was conducted on 30 seventh-grade students at Junior High School. This stage was carried out to evaluate the weaknesses and shortcomings of the developed product before proceeding to the next stage. The analyses to be conducted at this stage include item validity testing, reliability testing, item discrimination testing, and difficulty level testing. The following are the results of the small-group pilot test analysis. Item validity testing was conducted to determine the accuracy of the test items in measuring scientific creativity. Item validity results were obtained from statistical analysis performed on the students’ response data. The statistical analysis used was Pearson’s product-moment correlation. Each test item was considered valid if the calculated r -value was greater than the table r -value. Conversely, an item is deemed invalid if the calculated r is smaller than the table r (Azizah et al., 2023). Table 8 presents the results of the item validity test conducted in the small group pilot study.

Table 8 Results of the item validity test

Question No.	$r_{\text{calculated}}$	r_{table}	Category
1	0.41	0.34	Valid
2	0.64	0.34	Valid
3	0.55	0.34	Valid
4	0.41	0.34	Valid
5	0.40	0.34	Valid
6	0.40	0.34	Valid
7	0.33	0.34	Invalid

Based on Table 8, the validity test results show that the items with a calculated r value greater than the table r value are items 1, 2, 3, 4, 5, and 6. The item with a calculated r value less than the table r value is item 7, which falls into the category of invalid items. One of the factors causing invalid items is improper item construction and unclear instructions (Syarifudin, 2020). Improper item construction and unclear instructions can lead to misunderstandings among students when interpreting the items, resulting in answers that do not meet expectations. The invalid test item is question number 7. According to Azizah et al. (2022), questions deemed invalid cannot be used as is and must be revised so they can be reused. Revisions were made to the wording used in the question to make it easier for students to understand the instructions.

Reliability testing was conducted to measure the consistency and stability of item assessment results (Anggraini et al., 2022). An assessment instrument is considered reliable if the results are similar when administered repeatedly to the same individuals at different times or to different individuals at the same time (Noor & Fauzi, 2025). The formula used in item reliability testing is Cronbach's alpha. If the Cronbach's alpha result is high, the item is considered reliable. Table 9 presents the results of the reliability test from the small-group pilot study.

Table 9 Reliability test results

Item	Number of Students	Cronbach's Alpha	Criteria
7	30	0.56	Moderate reliable

The results of the reliability test indicate that the reliability coefficient of the scientific creativity assessment instrument is 0.56, which falls into the "moderate reliable" category. This figure suggests that the environment-based scientific creativity assessment instrument can be pilot tested, though it still requires minor revisions to the item construction.

The discriminative power of a test aims to identify differences in students' abilities (Tobin, 2024). The discriminative power index of a test item can serve as an indicator of the item's quality in distinguishing between students. The higher the index of a test item, the better the item, as it has the ability to distinguish between students who have mastered the material and those who have not. Conversely, the lower the item discrimination index value, the poorer the item is because it has a low ability to distinguish between students who have understood the material and those who have not (Mulyani et al., 2024). The results of the item discrimination test are shown in Table 10.

Table 10 Results of the Item discrimination test

Discrimination Criteria	Item Number	Count
Excellent	-	-
Good	2	1
Fair	5	1
Poor	1,3,4,6,7	5

Based on the results of the discriminant index test conducted on the 7 items, it can be seen that 1 item, specifically item 2, falls into the "good" category, with discriminant power indices ranging from 0.30-0.39, and 1 item, specifically item 5, falls into the "fair" category, with discriminant power indices ranging from 0.20-0.29. The remaining 5 items, questions 1, 3, 4, 6, and 7, fall into the poor category, with discrimination indices ranging from 0.19 and above. This result indicates that most items were not yet able to optimally distinguish between students with high and low levels of scientific creativity. Several factors may explain this finding, including the limited number of participants in the small group trial, the relatively homogeneous abilities of the students, and their limited familiarity with scientific creativity tasks. In addition, the scoring rubric and item characteristics may require further refinement to better capture differences in students' scientific creativity

performance. A difficulty level test was conducted to determine the difficulty level of each item, classifying them as easy, moderate, or difficult. The results of the difficulty level test from the small-group pilot study are presented in Table 11.

Table 11 Difficulty level test results

Question No.	Difficulty Index	Category
1	0.74	Easy
2	0.67	Moderate
3	0.60	Moderate
4	0.69	Moderate
5	0.57	Moderate
6	0.33	Moderate
7	0.45	Moderate
Mean	0.57	Moderate

Based on the results of the difficulty level test conducted on 7 essay-type items, it can be seen that 1 item falls into the easy category with an index of 0.74, and 6 items fall into the moderate category with index values ranging from 0.30 to 0.70. The product revision stage aims to improve test items that did not meet the validity criteria during the item validity test. Based on the table of item validity test results from the small-group pilot test, there is 1 invalid item, namely, item number 7; therefore, revisions must be made to this item to ensure it is suitable for the field pilot test.

Product revision and field testing

The field trial phase was conducted after the assessment instrument was revised based on the small-group pilot test to ensure that the developed assessment instrument is valid and reliable according to the criteria for a good assessment instrument. Corrections were made to the wording used in question 7, which was classified as invalid, to make it easier for students to understand the instructions. The large-group pilot test was administered to a larger sample of students, specifically seventh-grade students (Classes A, B, and C) at Junior High School, totaling 105 students. The results of the field trial were analyzed using Microsoft Excel to determine the validity, reliability, discriminant power, and difficulty level of the environment-based scientific creativity assessment instrument.

The validity test in the field trial involved 105 students, so $n = 105$ and $r_{\text{table}} = 0.15$. The results of the validity test can be seen in Table 12.

Table 12 Validity test results

Item No.	$r_{\text{calculated}}$	r_{table}	Category
1	0.64	0.15	Valid
2	0.68	0.15	Valid
3	0.72	0.15	Valid
4	0.70	0.15	Valid
5	0.59	0.15	Valid
6	0.61	0.15	Valid
7	0.66	0.15	Valid

Based on Table 12, the results of the item validity analysis for the large group indicate that each item has a calculated r value greater than the table r value. Therefore, it can be concluded that all 7 items in the development study are valid and can be used as an assessment instrument for environment-based scientific creativity in science classes at junior high schools.

The reliability results of the assessment instrument can be obtained from the Cronbach's alpha calculation. The higher the Cronbach's alpha value, the more reliable the assessment instrument is, and vice versa. The reliability test results can be seen in Table 13.

Table 13 Reliability test results

Question item	Number of student	Cronbach's Alpha	Criteria
7	105	0.78	High reliable

The results of the reliability test showed a Cronbach's alpha value of 0.78, which falls into the high reliable category. The increase in Cronbach's alpha in the field test may be influenced by several factors, including the clarity of the test items, the number of students, the students' condition, and the testing process ([Rasyid et al., 2025](#)). During the small-group test, the testing time overlapped with break time, resulting in a less conducive atmosphere due to the noise from students who had finished their break, and the students taking the test struggled to focus on the questions. This value indicates that all test items are reliable and suitable for use as an assessment instrument for environment-based scientific creativity in junior high school science classes.

The discriminant validity of the test items indicates how well the questions distinguish between students' abilities. Discriminant validity is important to ensure that the questions used can accurately distinguish between students who understand the material and those who do not. Table 14 presents the results of the discriminant validity test for the test items in the large group.

Table 14 Results of the Item discrimination test

Category	Item Number	Count
Excellent	3, 7	2
Good	2, 5, 6	3
Fair	1, 4	2
Poor	-	-

Based on the results of the item discriminant analysis in the table above, there are two items in the "excellent" category (3 and 7), three items in the "good" category (2, 5 and 6), and two items in the "fair" category (1 and 4). The results of the discriminant validity analysis for items in the "excellent", "good" and "fair" categories are suitable for use in assessing environment-based scientific creativity. The items developed by [Istiyova \(2025\)](#), which possess a sufficient level of discriminative power, continue to be used because they still have the ability to distinguish between high-ability and low-ability students.

A difficulty level analysis was conducted to obtain data on questions with easy, moderate, and difficult levels. Table 15 presents the results of the difficulty level analysis in the large-scale test.

Table 15 Results of the difficulty level test

Question No.	Difficulty Index	Category
1	0.72	Easy
2	0.71	Easy
3	0.59	Moderate
4	0.72	Easy
5	0.51	Moderate
6	0.27	Difficult
7	0.21	Difficult
Mean	0.53	Moderate

Based on the results of the difficulty level analysis of the 7 items in the environment based scientific creativity assessment instrument, there are items categorized as easy, moderate, and difficult. The items categorized as easy are items 1, 2, and 4, which have difficulty indices between 0.70 and 0.90; items in the moderate category are questions 3 and 5, which have indices of 0.30–0.70, and items in the difficult category are questions 6

and 7. The difficulty levels of test items are considered balanced or proportional if the ratio is 25% for the “easy” category, 50% for the “moderate” category, and 25% for the “difficult” category ([Pradita et al., 2023](#)). This indicates that the difficulty levels of the items in the environment-based scientific creativity assessment instrument are not appropriate, and it can be said that the proportion of the difficulty levels of these items is not yet balanced.

The improvement in the validity, reliability, and difficulty indices between the pilot and field testing phases may be attributed to two factors. First, revisions were made to Item 7, which was identified as invalid during the pilot test. The item was revised by improving its wording and alignment with the intended indicator before being administered in the field test. Second, the field testing involved a larger and more heterogeneous sample, resulting in more stable psychometric estimates. Consequently, the instrument demonstrated better overall measurement properties during the field testing phase.

CONCLUSION

The conclusion of this study is that the environment-based assessment instrument for scientific creativity in junior high school science has been found to be valid and reliable. The data were analyzed using expert validity and empirical validity tests, reliability tests, discriminant validity tests, and difficulty level tests. Recommendations for future research include developing an assessment instrument with a larger number of items and conducting a broader study. However, this study has limitations in terms of the number of questions developed and the fact that the development process only reached the field-testing stage. Further research involving a wider variety of questions and extending the development process to the stage of widespread distribution would enrich these findings.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS STATEMENT

The contributions of each author are as follows: Conceptualization, SN and IW; methodology, FHM; data collection, SN and FHM; formal analysis, SN and FHM; writing preparation of the original draft, SN and IW; and reference collection, SN and IW.

AI DISCLOSURE STATEMENT

The authors utilized a number of digital tools to enhance efficiency in the writing process. Secondary literature searches were conducted via Google Scholar, while citations and the reference list were managed using Mendeley. Quantitative data analysis was performed using Microsoft Excel, and the entire writing process was carried out using Microsoft Word. To ensure the originality of the manuscript and avoid plagiarism, this article has been checked using Turnitin. After using this service, the authors have thoroughly reviewed and edited the content as needed and take full responsibility for the content of this publication. The authors declare that this research was designed, conducted, written, and edited without the use of artificial intelligence (AI) technology.

REFERENCES

- Adetya, S., & Yuliana, Y. V. (2025). Systematic literature review: Validity and reliability testing of lecturer performance measurement tools. *Jurnal Fokus Manajemen*, 5(2), 317-324.
- Akbar, A. R. M., Khairunnisa, A., Sari, I. P., Atsir, R. R., Gumelar, R. M., Budiargo, W. F., & Sukmawati, M. (2025). Hakikat pendidikan ipa. *Algoritma : Jurnal Matematika*,

- Ilmu Pengetahuan Alam, Kebumihan Dan Angkasa*, 3(1), 235–245. <https://doi.org/10.62383/algoritma.v3i1.395>
- Anggraini, F. D. P., Aprianti, A., Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran statistika menggunakan software spss untuk uji validitas dan reliabilitas. *Jurnal Basicedu*, 6(4), 6491–6504. <https://doi.org/10.31004/basicedu.v6i4.3206>
- Augustia, A. D., Agustia, C. N., Azzahra, D., Hakim, L., & Pratiwi, V. (2025). Analisis validitas dan reliabilitas soal pilihan ganda dengan menggunakan software anates pada mata pelajaran perpajakan. *AKADEMIK: Jurnal Mahasiswa Ekonomi & Bisnis*, 5(1), 250-265. <https://doi.org/10.37481/jmeb.v5i1.1165>
- Azizah, I., Wahyuni, S., & Budiarso, A. S. (2023). Pengembangan instrumen penilaian berbasis literasi sains menggunakan quizizz untuk mengukur hots pada pembelajaran ipa siswa smp. *Paedagogia: Jurnal kajian Penelitian dan Pengembangan*, 14(2), 121–132. <https://doi.org/10.31764>
- Azizah, Z., Budijastuti, W., (2022). Development of assessment instrument to measure science literacy skill in the submatery of the human blood circulation. *Berkala Ilmiah Pendidikan Biologi*, 11(1), 89-97. <https://doi.org/10.26740/bioedu.v11n1.p89-97>
- Dahlan, A. Z., Shidiq, M., Anwar, Z., Syihabul, M., Al Haqiqy, I., & Qomari, N. (2024). Analysis of different power levels, difficulty, and effectiveness of distracting final semester assessment questions (pas) for arabic language class viii mts irsyadul athfal. *Journal of Islamic Education Studies*, 9(3), 398-412. <https://doi.org/10.58788/alwijdn.v9i3.4366>
- Fadllan, A., Hartono, H., Susilo, S., & Saptono, S. (2022). Pengembangan tes kreativitas saintifik fluida statis berbasis scientific creativity structure model. *Prosiding Seminar Nasional Pascasarjana*, 72-78. <http://pps.unnes.ac.id/prodi/prosiding-pascasarjana-unnes/72>
- Faiz, A., Putra, N. P., 7 Nugraha, F. (2022). Memahami makna tes, pengukuran (measurement), penilaian (assesment), dan evaluasi (evaluation) dalam pendidikan. *Jurnal Education and Development*, 10(3), 492-495.
- Fitriani, N. (2021). Analisis tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal pelatihan kewaspadaan kegawatdaruratan maternal dan neonatal. *Paedogaria*, 12(2), 199–205. <https://doi.org/10.31764>
- Genarsih, T., Tisngati, U. (2024). Belajar statistika: Konsep dasar dan pengantar statistika penelitian. Yogyakarta: Zahir Publishing.
- Hamidah, M. H., & Wulandari, S. S. (2021). Pengembangan instrumen penilaian berbasis hots menggunakan aplikasi “quizizz” This work is licensed under a creative commons attribution-sharealike 4.0 International license. *Kajian Ilmu Administrasi*, 18(1), 105–124. <https://doi.org/10.21831/efisiensi.v18i1.36997>
- Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, 24(4), 389–403. <https://doi.org/10.1080/09500690110098912>
- Indiana, S., Amaliyah, N., & Hartini, T. I. (2024). Hubungan antara kemampuan berpikir kritis dan kemampuan berpikir kreatif dengan penguasaan konsep dasar ipa pada siswa kelas v di SDN gugus 2 kecamatan cipayung kota depok. *Pedagogi: Jurnal Penelitian Pendidikan*, 11(1), 86-104. <https://doi.org/10.25134/pedagogi.v11i1.8507>
- Istiyova, L. R. (2025). Analisis butir soal pengukuran kemampuan berpikir kritis materi pencemaran lingkungan melalui teori tes klasik dan Rasch model. *Bio-Pedagogi*, 14(1), 36. <https://doi.org/10.20961/bio-pedagogi.v14i1.88792>
- Jannah, F., Radiansyah, R., Sari, R., Kurniawan, W., Aisyah, S., Wardini, S., & Fahlevi, R. (2022). Pembelajaran hots berbasis pendekatan lingkungan di sekolah dasar.

- Primary: *Jurnal Pendidikan Guru Sekolah Dasar*, 11(1), 189.
<https://doi.org/10.33578/jpkip.v11i1.8533>
- Judijanto, L., Haryani, H., Sari, N., Pranata, A., Mutoharoh, M., Lumbu, A., ... & Wiradika, I. N. I. (2025). *Assessment, testing dan evaluasi*. PT. Sonpedia Publishing Indonesia.
- Juliastari, J., Artayasa, I. P., & Merta, I. W. (2022). Pengaruh pembelajaran berbasis proyek dengan pendekatan sains teknologi masyarakat terhadap kreativitas ilmiah siswa. *Jurnal Ilmiah Profesi Pendidikan*, 7(2), 337–343.
<https://doi.org/10.29303/jipp.v7i2.444>
- Kareem, A. (2022). Higher-order thinking skills and scientific attitudes components as predictors of scientific creativity among preservice biology teachers. *International Journal of Progressive Education*, 18(4), 21–30.
<https://doi.org/10.29329/ijpe.2022.459.2>
- Moto, A., Musyarofah, L., & Taufik, ani S. (2022). Developing item analysis of teacher-made test for summative assessment of seventh grade of smpn 8 komodo in academic year 2020/2021. *Budapest International Research and Critics Institute (BIRCI-Journal)*. <https://doi.org/10.33258/birci.v5i1.4237>
- Mulyani, S. I., Syahidah, S. I., Oktavia, I., Zuhri, S., & Hidayat, W. (2024). Analisis kualitas soal evaluasi pada buku ajar pai dan budi pekerti smkn 3 kota serang. *ALFIHRIS: Jurnal Inspirasi Pendidikan*, 3(1), 119–131.
<https://doi.org/10.59246/alfihris.v3i1.1163>
- Nasruddin, N., Herlina, h., & Madil, S. (2021). Analysis of difficulty levels and discriminating power of items made by junior high school mathematics teachers. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 4(4), 11934-11942. <https://doi.org/10.33258/birci.v4i4.3258>
- Noor, N. H. H., & Fuzi, A. M. (2025). Assessment of validity, reliability, and normality in quantitative study: A survey instrument analysis with ibm spss. In *8th ASNet International Multidisciplinary Academic Conference* (pp. 27-43).
- Pradita, E., Megawanti, P., & Yulianingsih, Y. (2023). Analisis tingkat kesukaran, daya pembeda, dan fungsi distraktor pts matematika smpn jakarta. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan matematika*, 3(1), 109-118.
- Rasyid, A., Hartanto, R., & Priyanto, C. (2025). Reliabilitas alat ukur, jenis-jenis, cara pengukuran, dan faktor-faktor yang mempengaruhinya. *TSAQOFAH*, 6(2), 1285–1294.
<https://doi.org/10.58578/tsaqofah.v6i2.8511>
- Sidek, R., Halim, L., Buang, N. A., & Mohamad Arsad, N. (2020). Fostering scientific creativity in teaching and learning science in schools: A systematic review. *Jurnal Penelitian Dan Pembelajaran IPA*, 6(1), 13. <https://doi.org/10.30870/jppi.v6i1.7149>
- Sonra, M. F. R., & Suriani, A. (2025). Analisis faktor rendahnya pemahaman siswa tentang kerusakan lingkungan di sekolah dasar negeri 12 padang koto gadang. *Jurnal Cakrawala Pendidikan dan Biologi*, 2(2), 126-142.
<https://doi.org/10.61132/jucapenbi.v2i2.316>
- Sudarti, D. O. (2020). Mengembangkan kreativitas aptitude anak dengan strategi habituasi dalam keluarga. *Jurnal Al-azhar Indonesia Seri Humainaiora*, 5(3), 117-127.
<http://dx.doi.org/10.36722/sh.v5i3.385>
- Sugiono, S., Noerdjannah, N., & Wahyu, A. (2020). Uji validitas dan reliabilitas dan alat ukur SG posture evaluation. *Jurnal Keterampilan Fisik*, 5(1), 1-61.
<https://doi.org/10.37341/jkf.v5i1.167>
- Syaifudin, S. (2020). Validitas dan reliabilitas instrumen penilaian pada mata pelajaran bahasa arab. *Cross-Border: Jurnal Kajian Perbatasan antarnegara*. 3(2), 106-118.
- Tiana, A., & Wahyuni, S. (2025). Analisis strategi pembelajaran abad 21 untuk meningkatkan kreativitas ilmiah siswa SMP: tinjauan literatur sistematis. *Kalam*

- Cendekia: *Jurnal Ilmiah*, 13(2), 1440-1450.
<https://doi.org/10.20961/jkc.v13i2.103547>
- Tobin, M. A. (2024). Guide to item analysis. Schreyer Institute for Teaching Excellence, <https://creativecommons.org/licenses/by-nc-sa/4.0/>
- Wicaksono, I., & Erlina, N. (2024). The effect of virtual science teaching model on scientific creativity and learning outcomes. *Jurnal Pendidikan MIPA*, 25(1), 440–452.
<https://doi.org/10.23960/jpmipa/v25i1.pp440-452>
- Wicaksono, I., Supeno, & Budiarmo, A. S. (2020). Validity and practicality of the biotechnology series learning model to concept mastery and scientific creativity. *International Journal of Instruction*, 13(3), 157–170.
<https://doi.org/10.29333/iji.2020.13311a>
- Widyasari, C. (2023). *Kreativitas dan keberbakatan*. Muhammadiyah University Press.
- Xu, S., Reiss, M. J., & Lodge, W. (2025). Comprehensive scientific creativity assessment (c-sca): A new approach for measuring scientific creativity in secondary school students. *International Journal of Science and Mathematics Education*, 23(2), 293–319. <https://doi.org/10.1007/s10763-024-10469-z>

AUTHOR PROFILES



Full Name: Syinta Nur
 Email: syintanur04@gmail.com
 Affiliation: Science Education, University of Jember



Full Name: Dr. Iwan Wicaksono, S.Pd., M.Pd., M. C. E.
 Email: iwanwicaksono@fkpunej.ac.id
 Expertise: Science Education
 Affiliation: Science Education, University of Jember
<https://scholar.google.co.id/citations?user=FdCnaLoAAAAJ&hl=id>
<https://sinta.kemdiktisaintek.go.id/authors/profile/6004343>
<https://orcid.org/0000-0003-0717-1577>
<https://www.scopus.com/authid/detail.uri?authorId=57191155659>



Full Name: Dr. Fadiatul Hasinah Muhammad
 Email: fadiatul@uitm.edu.my
 Expertise: Advanced Materials, Solid State Ionics
 Affiliation: Centre of Foundation Studies
<https://shorturl.at/JVxCy>
<https://orcid.org/0000-0003-2111-6865>
<https://www.scopus.com/authid/detail.uri?authorId=35069999700>