14 (2) (2025) 290-300

Unnes Science Education Journal
Accredited Sinta 2

https://journal.unnes.ac.id/journals/usej

Development and Validation of a Research Skills Instrument for Students in the Context of Environmental Change to Support the Clean Water and Sanitation Goal

Ghurrotul Bariroh, Siti Sriyati✉, Rini Solihat

DOI: http://dx.doi.org/10.15294/usej.v13i1.22730

Universitas Pendidikan Indonesia, Indonesia

Article Info

Submitted 2025-06-17
Revised 2025-07-10
Accepted 2025-08-30

Keywords

Development, Education for Sustainable, Environmental Change, Instrument, Research Skills

Abstract

Research skills are an essential competencies in biology education, particularly in understanding environmental changes topic. One crucial aspect of this topic is the effort to maintain the quality of natural resources, including clean water, which aligns with Sustainable Development Goal (SDG) 6: Clean Water and Sanitation. However, there is currently no specific instrument designed to assess students' research skills, especially in the context of environmental change related to clean water and sanitation. Therefore, this study aims to develop a multiple-choice research skills assessment instrument. The development process followed the 4D model, which includes the stages of Define, Design, Develop, and Disseminate. Data were analyzed using descriptive statistics to evaluate the validity, reliability, and feasibility of the instrument, and the results were interpreted to determine the instrument's effectiveness in assessing students' research skills. Expert validation results indicate that the instrument has a feasibility score of 82.55%. A field trial involving 63 students revealed that out of 26 test items, 13 were accepted, 5 required revision and 8 were rejected based on validity, reliability, difficulty level, and discrimination index analyses. Overall, the instrument is valid, reliable, and effective in measuring students' research skills. Future research is recommended to develop more contextually relevant instruments that better reflect students' research skills in addressing real-world challenges. Additionally, further validation in broader educational settings is necessary to ensure the instrument's effectiveness across various instructional methods.

How to Cite

Bariroh, G., Sriyati, S., & Solihat, R. (2025). Development and Validation of a Research Skills Instrument for Students in the Context of Environmental Change to Support the Clean Water and Sanitation Goal. *Unnes Science Education Journal*, 14(2), 290-300.

✉ Correspondence Author:
E-mail: sriyati@upi.edu

p-ISSN 2252-6617
e-ISSN 2502-6232

INTRODUCTION

In the era of 21st century learning, students are no longer solely focused on mastering content but are also encouraged to develop higher-order thinking and applied skills, including research skills (Barus et al., 2020; Trilling & Fadel, 2009). As a crucial component of 21st century learning skills, research skills enable students to critically investigate problems, generate and evaluate relevant data, and also formulate and test theories and hypotheses. This process enhances students' understanding of subject matter through research-based activities (Webb et al., 2011; Willison & O'Regan, 2007). Engaging in research requires students to undergo a series of investigative processes, data analysis, and evidence-based decision-making. Consequently, research skills help students explore information in depth, comprehend complex global issues, and develop innovative solutions applicable to real-world contexts. Therefore, Mutlu (2020) emphasizes that research skills should be instilled in students to equip them for the challenges of global dynamics, as they indirectly assist students in addressing real-life problems in their surroundings.

Research skills are essential competencies in science education, particularly in biology. The primary goal of biology education is to enhance students' learning experiences and facilitate the development of scientific knowledge through inquiry-based activities (Natale et al., 2020; Spence, 2020). Therefore, research skills must be cultivated at the secondary education level, as they provide students with a critical foundation for scientific exploration (Maddens et al., 2021). To effectively develop research skills, educational programs should be designed to equip students with the ability to navigate the challenges and opportunities of sustainable development. As a demonstration of the Indonesian government's commitment to fostering research skills, these competencies have been integrated into the Graduate Competency Standards at the General Secondary Education level, as stated in the Regulation of the Minister of Education, Culture, Research, and Technology of the Republic of Indonesia Number 5 of 2022, Article 9. Additionally, this commitment is reflected in the competency framework of the Merdeka Curriculum, which explicitly outlines key scientific process skills, including observation, inquiry, planning and conducting investigations, data processing and analysis, evaluation, reflection, and effective communication of findings.

Based on the results of Noorlathifah's analysis (2024), 81% of learning outcomes Phase E and F of biology subjects require research skills to meet these learning outcomes. These learning outcomes include the classification of living things; viruses, bacteria, and fungi; ecosystems and interactions between components and influencing factors; biotechnology; climate change; cell structure, cell division, and membrane transport; metabolism and protein synthesis, growth and development, evolutionary theory and linking it to biodiversity; and the relationship of organ structure to organ systems and their functions. The learning material on the impact of pollution that endangers water quality and aquatic ecosystems is one of the learning topics suggested by UNESCO (2017) in achieving the 6th SDG. In Indonesia, the material is integrated in the climate change learning outcomes, which at the end of Phase E and F students have the ability to understand climate change, so that they are responsive and can play an active role in solving problems on local and global issues. Therefore, to achieve these learning objectives, students need to develop research skills that enable them to identify various causes of pollution, utilize technology to analyze pollution levels, and design and carry out experiments as alternative solutions.

UNICEF and WHO (2020) recognize water pollution as a problematic issue that has a major impact on multi-sectors such as economic, educational, environmental, health and social development. One of the consequences of water pollution is the decreased availability of clean water. The existence of water not only supports the existence of living organisms, but is also a key factor in maintaining the global ecosystem and climate balance (Kilic, 2020). The decrease in the availability of clean water due to water pollution is further exacerbated by the rapid growth of the population and economic development which also drives the demand for water to continue to increase, making the water supply gap more complicated to overcome (Baggio et al., 2021; Boretti & Rosa, 2019). Amidst competing water demands in various sectors, the availability of clean water for environmental needs is critical to ensure ecosystem sustainability. Therefore, ensuring the availability of clean water is the responsibility of the entire community, which is also an integral part of the 6th Sustainable Development Goal (SDGs), namely Clean Water and Sanitation. The 6th goal of the SDGs emphasizes the importance of universal access to clean water and safe sanitation to achieve good health, reduce poverty, and improve the quality of human life globally (United Nations, 2025). This goal underscores

the need to protect and restore vulnerable water ecosystems and improve water management efficiency at all levels.

To determine the level of students' research skills, a valid and reliable instrument is needed as a measuring tool that can provide objective picture of students' ability to conduct scientific investigations. In the Research Skill Development Framework developed by Willison and O'Regan (2007), research skills instruments must be able to assess various aspects ranging from initiating and clarifying problems, finding and generating data, evaluating and reflecting on data, organizing and managing data, analyzing and synthesizing data, communicating and applying results. Some previous studies have developed research skills instruments, such as Symsons et al. (2017) developed a questionnaire focusing on students' research motivation and awareness, Molina (2019) developed an instrument for high school students in general science learning, Ilah et al. (2021) developed student research skills test for authentic research programs, Daryanes et al. (2024) developed research skills instrument for prospective biology teachers. However, these studies have not specifically measured students' research skills, especially in context of environmental change as effort achieve Clean Water and Sanitation (6th SDGs).

Based on this research gap, this study aims to develop research skills instruments that are specific, comprehensive, and in accordance with the demands of 21st century learning, especially in the context of environmental changes oriented towards Clean Water and Sanitation. The research skills instrument developed adopted the Research Skill Development Framework level-2 (Bounded Research) by Willison and O'Regan (2007). The use of level 2 adjusts the characteristics of students at the high school level where research activities are still limited to teacher guidance (Hendrawati, 2022; Sutia, 2018). With a structured instrument based on high validity and reliability, educators can objectively measure the development of students' research skills, design more effective learning interventions, and evaluate the effectiveness of research-based teaching methods. The development of this instrument is expected to not only improve students' understanding of scientific concepts, but also equip them with critical thinking and problem-solving skills that are essential in facing the challenges of the global environment.

METHOD

This study used a research and development approach to develop a student research skills test instrument in environmental change content oriented towards achieving Clean Water and Sanitation (6th SDG). The method used refers to the 4D development model by Thiagarajan et al. (1974) which consists of four stages: define, design, develop, and disseminate. The illustration of the 4D model can be seen in Figure 1. The data collection in this study is in the form of expert validation feedback and data from field trials.

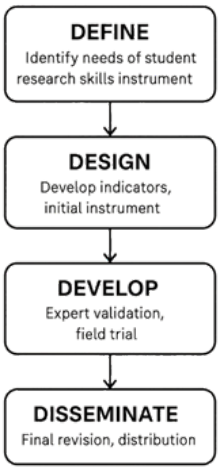

Figure 1. Flowchart of the 4D Research and Development Model for Developing Research Skill Instruments

Development Stage

At the define stage, the need for student research skills instruments on environmental change material oriented towards Clean Water and Sanitation (6th SDGs) was identified through literature review and empirical analysis. A literature review was conducted to examine students' research skills in the context of environmental change, by referring to international journals. Empirical analysis was conducted through interviews with biology teachers to identify the types of assessments that have been used to measure students' research skills, as well as challenges in implementing them in the classroom.

At the design stage, the instrument was developed by compiling indicators of research skills to be measured based on the Research Skill Development Framework level 2 (Bounded Research) by Willison and O'Regan (2007). The design of the instrument includes the question format, instructions, answer key and scoring method. The instrument was developed in the form of a multiple choice test consisting of 6 indicators and then broken down into 12 sub-indicators, so that the initial instrument produced 26 multiple choice questions as in Table 1.

Table 1. Blueprint of Research Skills Test Instrument

Facet	No	Sub-indicator	Number of Item
Embark & Clarify	1	Identifying research questions from a closed inquiry context.	2
	2	Formulating terms, conditions, expectations, or research hypotheses from provided options.	2
Find & Generate	3	Collecting relevant data from predetermined sources when the accuracy of previous data is uncertain.	2
	4	Recording accurate data from predetermined sources when the accuracy of previous data is uncertain.	2
Evaluate & Reflect	5	Evaluating data using selected criteria provided by the teacher to determine data credibility.	2
	6	Reflecting on the research process using selected criteria provided by the teacher.	2
Organize & Manage	7	Organizing data using a given structural framework.	2
	8	Managing the research process with possible alternative pathways.	2
Analyze & Synthesize	9	Interpreting multiple data sources in a standard format.	2
	10	Synthesizing multiple data sources to integrate with existing knowledge in a standard format.	2
Communicate & Apply	11	Using a specified genre to develop and demonstrate understanding for a predetermined audience.	2
	12	Applying acquired knowledge to a similar context.	2
	13	Following research guidelines related ethnic, social, cultural issues.	2

At the develop stage, the research skills test instrument was prepared in the form of multiple choice questions with a context-based assessment approach. Expert validation was conducted by two biology education lecturers and a high school biology teacher covering aspects of construction, content, and language to review content validity and ensure the relevance of the instrument in learning. The data that has been obtained from the results of expert validation is then calculated the percentage using the following formula.

$$Percentage\ of\ feasibility = \frac{score\ obtained}{maximum\ score} \text{ x } 100$$

(Riduwan, 2013)

The percentage obtained is then interpreted in the criteria in Table 2.

Table 2. Interpretation of expert validation scores

No	Percentage (%)	Category
1	81.25 – 100.00	Highly Feasible
2	62.50 – 81.24	Feasible
3	43.75 – 62.49	Not Feasible
4	25.00 – 43.73	Highly Not Feasible

(Sugiyono, 2019)

After validation, the field trial was conducted on 63 students in grade XI of senior high school. The test subjects were chosen because they had studied environmental change material and had experience in relevant laboratory practices, thus enabling them to engage in research-based experimental activities.

In the last stage of disseminate, after going through the final revision process, the instrument that has been validated and proven reliable is distributed to biology teachers to be applied more widely in measuring students' research skills on environmental change material.

Analysis of Field Trial Results

The data from the trial were analyzed using the product moment test of SPSS version 27 to test the validity of the instrument. Meanwhile, the reliability test, difficulty level, and differentiating power of the questions were analyzed using ANATES V4 to ensure the quality of the items and the internal consistency of the instrument. The data from the item analysis that has been carried out is then categorized according to the item decision-making criteria table as in Table 3. The instrument items that fall into the "Accept" and "Revision" categories can be used. Items in the "Accept" category meet all criteria and can be used directly without modification. Meanwhile, items in the "Revision" category may be used after improvements are made based on their validity, discrimination index, or difficulty level.

RESULT AND DISCUSSION

Results

1. Expert Feasibility Test Result

The development of student research skills instruments was carried out through a series of stages, namely define, design, develop, and dissa-

Table 3. Criteria for Making Question Item Decisions

Category	Criteria
Accept	Validity ≥ 0.40 Discrimination index≥ 0.40 Level of difficulty 0.25 ≤ P ≤ 0.80
Revision	Discrimination index ≥ 0.40; level of difficulty P < 0.25 or P > 0.80; but validity ≥ 0.40 Discrimination index < 0.40; level of difficulty 0.25 ≤ P ≤ 0.80; but validity ≥ 0.40 Discrimination index ≥ 0.40; level of difficulty 0.25 ≤ P ≤ 0.80; but validity between 0.20 – 0.40
Reject	Discrimination index = 0.40 and level of difficulty P < 0.25 or P > 0.80 Validity < 0.20 Discrimination index < 0.40 and validity < 0.40

(Zainul, 2002)

minate. At the develop stage, after the instrument was completed, it was validated by experts to ensure the quality, suitability, and feasibility of the instrument in measuring students' research skills. This validation includes three main aspects, namely construct, content, and language (Table 4).

Table 4. Results of Expert Validation of Student Research Skills Instruments

No	Aspect	Average Expert score			%	Category
		1	2	3		
1	Construct	88	83	86	85.66	Very Feasible
2	Content	82	81	81	81.33	Very Feasible
3	Language	78	81	83	80.66	Feasible
Average					82.55	Very Feasible

The expert validation was conducted to assess the feasibility of the student research skills instrument in terms of construct, content, and language. Based on expert validation results, the student research skills test instrument met the feasibility standards and was rated as highly feasible. The highest score was obtained for the construct aspect, followed by the content aspect, which was also rated as highly feasible. However, the language aspect which scored lower than the other two aspects, indicates that there are still some items that need to be improved even though they are classified as feasible. An example of the form of improvement as suggested by the expert can be seen in Table 5.

Table 5. Example of Instrument Improvement Based on Expert Feedback

Sub-indicator	No	Initial version	Revised version
Recording accurate data from predetermined sources when the accuracy of previous data is uncertain.	4	Your group is conducting a phytoremediation experiment using lotus in reducing DO (Dissolved Oxygen) levels of coffee wastewater that pollutes the surrounding environment. During the data collection process, the DO levels are higher than expected and data from previous reference sources. What should you do to record the right data? A. It remains to record the data as it is without matching it with other sources because the data is taken from direct observation. B. Ignore the data because it does not match with reliable references. C. Recording the measurement data care fully, then comparing it with other research from reliable sources. D. Using data from other sources so that the research results are in accordance with existing references. E. Not recording the measurement results until getting the same data as previous research.	Your group is conducting a phytoremediation experiment using lotus to reduce dissolved oxygen (DO) levels in coffee wastewater that pollutes the river around the school. After the phytoremediation process, the measurement results show that the DO levels are higher than expected and different from the results of previous studies. What steps should you take to ensure the data recorded is correct? A. Record the measurement results as they are without comparing them with other sources because they are obtained from direct observation. B. Ignoring the data because it is not in accordance with existing references. C. Recording the measurement data carefully, then comparing it with other research from trusted sources. D. Using data from other sources so that the research results are in accordance with existing references. E. Not recording the measurement results until getting the same data as previous research.

One of the revisions was made to the Find & Generate facet items with the sub-indicator of recording the right data from a predetermined source when the previous data was not clear. In

the initial version, the sentences in the question were still less specific, such as the use of the term "DO (Dissolved Oxygen) levels of coffee liquid waste polluting the surrounding environment," which could cause confusion because it did not clearly explain where the pollution occurred. In addition, the phrase "because the data is taken from direct observation" in answer option A also has potential to cause multiple interpretations. Therefore, in the revised version, context of the question is clarified by mentioning that the research was conducted in the river around the school. Technical terms such as "reducing dissolved oxygen (DO) levels" are also made more explicit so as not to cause misunderstanding for students. In addition, in the question, phrase "the data collection process showed that DO levels were higher than expected" was changed to "measurement results showed that DO levels were higher than expected and different from the results of previous studies". This change aims to avoid ambiguity in data interpretation as well as ensure that students understand difference between actual measurements and previous references.

2. Filed Trial Result

The field trial was conducted to evaluate the quality of the test instrument items before being used to assess students' research skills in learning. This analysis encompassed several key aspects, including validity, reliability, discrimination index, and difficulty level. The validity test in this study aimed to ensure that each test item genuinely assessed research skills in accordance with the predetermined indicators. The interpretation of the validity test in this study refers to Arikunto (2016) who categorizes instrument validity criteria into very high, high, moderate, low, and very low. The results of the validity test for the student research skills test instrument are presented in Table 6.

Table 6. Validity Analysis of the Student Research Skills Test Instrument

Range	Interpretation	(%)	Question Number
0.80 – 1.00	Very High	8	10, 20
0.60 – 0.79	High	11	7, 11, 13
0.40 – 0.59	Moderate	46	1, 2, 3, 4, 6, 8, 9, 12, 14, 21, 22
0.20 – 0.39	Low	4	18
0.00 – 0.19	Very Low	31	5, 15, 17, 19, 23, 24, 25, 26

The validity analysis of the test items indicates that 8% of the items fall into the high category and 11% into the very high category, showing that these items effectively measure the intended research skills indicators. Furthermore, 46% of the items are in the moderate category, meaning they are still acceptable for use. In contrast, 4% of the items are classified as low and 31% as very low, indicating that these items do not sufficiently represent the aspects of research skills being measured and therefore require substantial revision or replacement.

The reliability test analysis using ANATES V4 yielded a score of 0.83. According to Arikunto's (2016) interpretation, this value falls within the very high category, indicating that the student research skills assessment instrument has a very high level of reliability. While the interpretation of difficulty level in this study also refers to Arikunto (2016), who categorizes question difficulty criteria into difficult, moderate, and easy. The results of the difficulty level analysis for the student research skills test instrument are presented in Table 7.

Table 7. Difficulty Level Analysis of the Student Research Skills Test Instrument

Range	Interpretation	(%)	Question Number
0.00 – 0.29	Difficult	35	5, 7, 10, 19, 20, 23, 24, 25, 26
0.30 – 0.69	Moderate	30	2, 6, 8, 9, 14, 18, 21, 22
0.70 – 1.00	Easy	35	1, 3, 4, 11, 12, 13, 15, 16, 17
0.20 – 0.39	Low	4	18
0.00 – 0.19	Very Low	31	5, 15, 17, 19, 23, 24, 25, 26

Based on the difficulty level analysis, 35% of the questions fall into the difficult category, 30% into the moderate category, and 35% into the easy category. This distribution shows that the proportion of questions in the instrument is relatively balanced and aligns with recommended standards. The percentages of difficult and easy questions are equal (35%), while the moderate category accounts for 30%, indicating that most test items are of appropriate difficulty and can be used without major revisions.

The discrimination index analysis was conducted to assess the extent to which a test item can differentiate between high- and low-ability students. The interpretation of the discrimination

index in this instrument follows Arikunto (2016), who classifies test item discrimination criteria into poor, fair, good, and excellent. The results of the discrimination index analysis for the student research skills test instrument are presented in Table 8.

Table 8. Analysis of the Discrimination Index of the Research Skills Test Instrument

Range	Interpretation	(%)	Question Number
0.00 – 0.19	Poor	35	4, 5, 15, 17, 19, 23, 24, 25, 26
0.20 – 0.39	Fair	4	12
0.40 – 0.69	Good	42	1, 2, 3, 6, 8, 9, 14, 16, 18, 21, 22
0.70 – 100	Excellent	19	7, 10, 11, 13, 20
0.00 – 0.19	Very Low	31	5, 15, 17, 19, 23, 24, 25, 26

Based on the discrimination index analysis, 42% of the questions fall into the "Good" category, while 19% are classified as "Excellent," indicating that the majority of the test items effectively differentiate between students with high and low levels of understanding. Additionally, 4% of the questions fall into the "Fair" category, suggesting that these items are less optimal in distinguishing students' abilities. Meanwhile, 35% of the questions are categorized as "Poor," meaning that these items fail to adequately differentiate students based on their understanding and therefore require substantial revision or replacement.

Discussion

The results of expert validation indicate that the student research skills test instrument meets the eligibility standards with a "highly feasible" category, meaning that this instrument can be used as a valid and reliable tool to evaluate students' research skills. The construct aspect assesses the alignment of each test item with the predetermined research skills indicators. This aspect received the highest score, indicating that the instrument effectively represents research skills through items that are in accordance with the established indicators. The clear and logical structure of the questions allows students to optimally demonstrate their research abilities, from data collection to the communication of research findings. Therefore, this instrument not only assesses theoretical understanding but also comprehensively reflects research skills in various research-based learning contexts.

The content aspect focuses on the alignment of the material in the test items with the curriculum and relevant scientific substance. Each item is analyzed to ensure that the material being tested is accurate, aligned with the concept of environmental change oriented toward the Clean Water and Sanitation goal, and free from misconceptions. The content aspect received a "highly feasible" rating, further reinforcing the instrument's alignment with the curriculum and relevant learning materials. The developed test items reflect contextual research scenarios, ensuring that students not only answer questions but also learn to apply research principles in authentic situations. This consistency is essential, as research skills are not only conceptual but also require applied understanding that can be cultivated through practice aligned with scientific standards (Subekti et al., 2018). With high content validity, this instrument serves as an effective evaluation tool for measuring the extent to which students can develop their research skills in a science-based and exploratory learning environment.

The language aspect evaluates clarity, readability, and grammatical appropriateness so that the questions do not contain ambiguity. Language clarity and appropriate sentence structure are very important in test instruments to avoid ambiguity that can affect students' understanding of the questions (Barr et al., 2019). The language aspect received a lower score than the other two aspects. This supports Constantinou (2020), who stated that linguistic factors in assessment instruments often pose challenges in ensuring clear communication of questions. However, this does not indicate that the instrument is unfeasible; rather, it suggests that minor refinements in wording could further enhance students' comprehension. Language accuracy significantly impacts the instrument's validity, as ambiguity in question statements can lead to biased student interpretations, ultimately affecting test reliability (Crisan et al, 2023; Vigueras et al., 2023). Further analysis should consider whether students' difficulty in understanding questions stems from limited academic vocabulary, overly complex sentence structures, or technical terms that are not adequately explained within the question context. If language-related issues predominantly arise in high-cognitive-level questions, it is possible that these difficulties are not solely due to linguistic aspects but also due to how the instrument structures cognitive complexity into accessible text for

students with varying levels of scientific literacy.

Item analysis is a crucial step in ensuring the quality of an evaluation instrument used to assess student competencies. A high-quality test item must meet three main criteria: validity, reliability, discrimination index, and difficulty level (Zainul, 2002). According to Sugiyono (2019), a valid instrument is one that accurately measures what it is intended to measure. The validity test in this study aimed to ensure that each test item genuinely assessed research skills in accordance with the predetermined indicators. The analysis of item validity showed that 11% of the student research skills test items fell into the high and 8% very high categories. Additionally, 46% of the items were categorized as moderate, while 3% were classified as low, and 31% were classified as very low. This variation in validity levels suggests that, during the development of the test instrument, some items may not have effectively measured the intended competencies or exhibited inconsistencies in their construction. As noted by Muliana et al. (2020), content validity largely depends on the extent to which the test items align with the measured concept and students' understanding when responding to them.

Reliability refers to the extent to which an instrument produces consistent results when measurements are repeated under the same conditions (Sugiyono, 2019). A reliable instrument yields stable and dependable outcomes, which is essential for ensuring the accuracy and credibility of research data. The results of the reliability analysis indicate that the instrument possesses a high level of reliability. This finding confirms that the test items function consistently across administrations, thereby strengthening the trustworthiness of the instrument as a tool for assessing students' research skills.

The difficulty level aims to determine whether a test item is difficult or easy, which can also be interpreted as the probability of students providing the correct answer at a certain ability level (Quaigrain & Arhin, 2017). According to Sudjana (2017), the ideal composition of test items follows a normal curve distribution, with a ratio of 30% easy, 40% moderate, and 30% difficult questions. Test items with an optimal difficulty level are those with a moderate difficulty index (Mardapi, 2017). Furthermore, Arikunto (2018) explains that questions that are too difficult may cause students to feel discouraged and lose motivation, as they are perceived as beyond their capability. Conversely, questions that are too easy fail to encourage students to think critically and put in more effort to solve them. Based on the difficulty level analysis, 35% of the items fall into the difficult category, 30% are moderately difficult, and 35% are easy. This distribution indicates that the proportion of questions in the instrument is well-balanced and aligns with the recommended standards. Moderately difficult questions tend to strike a balance between challenging students and remaining solvable for varying levels of student ability (Julianingsih et al., 2014).

The item discrimination analysis was carried out to evaluate how effectively each test item distinguishes between students with higher and lower levels of ability. A good question is one that can only be answered correctly by high-achieving students, whereas a question that can be answered correctly by both high- and low-achieving students is less effective because it lacks discrimination power (Fatimah & Alfath, 2019). The higher the discrimination index of a test item, the better it is at distinguishing between students who have mastered the competencies and those who have not (Nasruddin et al., 2021). Based on the results of the discrimination index analysis, 42% of the test items fall into the "Good" category, 19% are classified as "Excellent," 4% are categorized as "Fair," and 35% are classified as "Poor." These results indicate that although the majority of items (61%) are effective in distinguishing between students with high and low levels of understanding, a considerable portion (35%) has poor discrimination power and requires substantial revision or replacement.

Based on the item analysis results regarding validity, discrimination index, and difficulty level, the test items were categorized into three main groups: accepted items, items requiring revision, and rejected items. This classification ensures that each test item used to assess students' research skills meets high-quality instrument criteria, enabling accurate and fair measurement of student competencies.

A total of 13 test items in this instrument were categorized as accepted without revision. These items include numbers 1, 2, 6, 7, 8, 9, 11, 13, 14, 16, 20, 21, and 22. These items met all criteria, with validity ≥ 0.40, discrimination index ≥ 0.40, and an ideal difficulty level range ($0.25 \leq P \leq 0.80$). High validity indicates that these questions align with learning indicators and effectively measure the expected cognitive aspects. Additionally, a good discrimination index ensures that the questions can effectively differentiate between high- and low-achieving students, allowing for a more objective evaluation of student learning outcomes. With a balanced difficulty level, these questions are neither too difficult nor too easy,

ensuring they accommodate students of all ability levels without introducing bias in assessment. Therefore, these questions can be used directly without modifications.

Five questions require revision due to failing to meet at least one criterion. Specifically, questions 3, 4, 10, 12, and 18 indicating the need for modification to improve their overall quality and alignment with the instrument's standards. Finally, eight questions were rejected because they did not meet the minimum criteria. These questions were numbers 5, 15, and 17, 19, 23, 24, 25, and 26 which had validity scores below 0.40 and low discrimination indices (< 0.40), making them unsuitable for inclusion in the test instrument. These items need to be replaced with new questions or completely revised to better align with measurement quality standards.

The developed instrument serves not only as a tool to measure students' research skills but also as a means to strengthen their competence in addressing environmental challenges. Through its application, students are guided to critically analyze issues related to water pollution, investigate potential solutions, and apply scientific approaches to environmental problem-solving. Such skills are essential for fostering awareness and action toward sustainable water management, thereby supporting the achievement of Sustainable Development Goal 6 (Clean Water and Sanitation) in the context of ongoing environmental change.

CONCLUSION

The development of a student research skills instrument in the context of environmental change oriented towards Clean Water and Sanitation has resulted in a valid and reliable measurement tool. This instrument not only assesses theoretical understanding but also requires students to apply research skills in more authentic situations. The high construct and content validity indicate that the instrument aligns with the measured research skill indicators, while its reliability ensures that the assessment results are consistent and trustworthy. However, linguistic factors and the contextual relevance to real-world research challenges remain aspects that need to be strengthened to make this assessment more inclusive and reflective of the complexities of real-world research. While the instrument has been categorized as highly feasible, this study has several limitations that need to be addressed in future research. The current version primarily focuses on conceptual aspects of research skills and has not yet fully captured students' abilities to apply these skills in authentic, real-world environmental contexts. Moreover, the test items lack broader contextual scenarios that could better represent the complexities of dynamic environmental issues. Future studies are recommended to refine and expand the instrument by incorporating more performance-based assessments and scenario-based tasks, allowing for a more comprehensive evaluation of students' ability to design, conduct, and evaluate research. Such improvements would enhance the instrument's effectiveness as a tool not only for measuring research skills but also for preparing students to actively contribute to achieving SDG 6 through innovative and sustainable scientific practices.

REFERENCES

Arikunto, S. (2016). *Prosedur Penelitian Suatu Pendekatan Praktik. (Edisi Revisi)*. Jakarta: PT. Rineka Cipta.

Arikunto, S. (2018a). *Prosedur Penelitian*. Jakarta: PT. Rineka Cipta

Baggio, G., Qadir, M. & Smakhtin, V. (2021). Freshwater availability status across countries for human and ecosystem needs. *Science of The Total Environment,* 792, https://doi.org/10.1016/j.scitotenv.2021.148230

Barr, C., Uccelli, P., & Galloway, E. (2019). Specifying the Academic Language Skills That Support Text Understanding in the Middle Grades: The Design and Validation of the Core Academic Language Skills Construct and Instrument. *Language Learning*, 69(4), 978–1021. https://doi.org/10.1111/LANG.12365.

Barus, C. S. A., Rosiqoh, & Suhendi, E. (2020). Identifying scientific critical thinking skills of high school students on the static fluid. *Journal of Physics: Conference Series*, 1521(2). https://doi.org/10.1088/1742- 6596/1521/2/022047

Boretti, A. & Rosa, L. (2019). Reassessing the projections of the World Water Development Report. *npj: Clean Water*, 2(15), 1-6. https://doi.org/10.1038/s41545-019-0039-9

Constantinou, F. (2020). Examination questions as a form of communication between the examiner and the examinee: a sociolinguistic perspective on assessment practice. *Cambridge Journal of Education*, 50, 711 - 728. https://doi.org/10.1080/0305764x.2020.1761293.

Crisan, I., Matei, A., Avram, D., Bunghez, C., & Erdodi, L. (2023). Full of Surprises: Performance Validity Testing in Examinees with Limited English Proficiency. *Psychological Injury and Law,* 16(2), 177-198, 177-198. https://doi.org/10.1007/s12207-023-09474-4.

Hendrawati, M. A. (2022). Meningkatkan Keterampilan Riset Siswa SMA Melalui Pembelajaran Berbasis Proyek pada Materi Pencemaran

Lingkungan. *Tesis*. Universitas Pendidikan Indonesia. https://reader-repository.upi.edu/index.php/display/file/76887/4/7

Julianingsih, E., Retnawati, H., & Arliani, E. (2024). Analysis of statistics test based on Yogyakarta culture for grade VIII with Anbuso 8.0. *The 5th International Seminar of Innovation in Mathematics and Mathematics Education (5th ISIMMED) and The 7th International Seminar on Science Education (7th ISSE): Technological Pedagogical Content Knowledge (TPACK) in The Current Trend of STEM Education*, 2622(1). https://doi.org/10.1063/5.0133856.

Kilic, Z. (2020). The importance of water and conscious use of water. *International Journal of Hydrology*, 4(5), 239-241. https://doi.org/10.15406/ijh.2020.04.00250

Madden, M. E., Baxter, M., Beauchamp, H., Bouchard, K., Habermas, D., Huff, M., Ladd, B., Pearon, J., & Plague, G. (2013). Rethinking STEM education: An interdisciplinary STEAM curriculum. *Procedia Computer Science,* 20, 541–546. https://doi.org/10.1016/j.procs.2013.09.316

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan (Edisi 2)*. Yogyakarta: Parama Publising.

Menteri Pendidikan Kebudayaan Republik Indonesia. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 5 Tahun 2022 Tentang Standar Kompetensi Lulusan pada Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah*. Indonesia.

Muliana, N., Pada, A., & Nurmaliah, C. (2020). Content validity of conation assessment. *Journal of Physics: Conference Series*, 1460. https://doi.org/10.1088/1742-6596/1460/1/012057.

Mutlu, A. (2020). Evaluation of Student Scientific Process Skills Through Reflective Worksheet in The Inquiry Based Learning Environments. *Reflective Practice*, 21(2), 271-286. https://doi.org/10.1080/14623943.2020.1736999

Nasruddin, Herlina, & 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

Natale, C., Mello, P., Trivelato, S., Marzin-Janvier, P., & Manzoni-De-Almeida, D. (2020). Evidence of Scientific Literacy Through Hybrid and Online Biology Inquiry-Based Learning Activities. *Higher Learning Research Communications*. 11, 33-49. https://doi.org/10.18870/HLRC.V11I0.1199.

Noorlathifa, S. (2024). Profil Keterampilan Riset yang Dibekalkan dalam Buku Teks Biologi Fase E dan F Kurikulum Merdeka. *Skripsi*. Universitas Pendidikan Indonesia. UPI Repository. https://repository.upi.edu/117848/

Nurlaelah, I., Widodo, A., Redjeki, S. & Rahman, T. (2021). Pengembangan Instrumen Tes Keterampilan Meneliti Pada Program Riset Autentik. *Quagga: Jurnal Pendidikan dan Biologi*, 13(2), 37-44. doi: 10.25134/quagga.v13i2.4232.

Quaigrain, K., & Arhin, A. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1). https://doi.org/10.1080/2331186X.2017.1301013.

Riduwan. 2013. *Dasar-dasar Statistik*. Bandung: Alfabeta.

Spence, P., Phillips, R., Mcallister, A., White, S., & Hollowell, G. (2020). Student-Scientist Curriculum: Integrating Inquiry-Based Research Experiences and Professional Development Activities into an Introductory Biology Laboratory Course†. *Journal of Microbiology & Biology Education*, 21. https://doi.org/10.1128/jmbe.v21i3.2225.

Subekti, H., Yuhanna, W. L., Susilo, H., Ibrohim & Suwono, H. (2018). Representation of Mutual Terms and Research Skills Towards Grade Point Average: Exploration Study. *Florea*, 5(1), 1-10. https://doi.org/10.25273/florea.v5i1.2500

Sudjana, N. (2017). *Penilaian Hasil Proses Belajar Mengajar.* Bandung: Remaja Rosdakarya.

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung : Alphabet.

Sutia, C. (2018). Membangun Keterampilan Riset Abad ke-21 Siswa Melalui Learning Management System Berbasis Google Classroom pada Pembelajaran Proyek Biologi. *Tesis*. Universitas Pendidikan Indonesia. https://reader-repository.upi.edu/index.php/display/file/40494

Symons, S. L., Colgoni, A., & Harvey, C. T. (2017). Student Perceptions of Staged Transfer to Independent Research Skills During a Four-year Honours Science Undergraduate Program. *In The Canadian Journal for the Scholarship of Teaching and Learning*, 8(1). http://ir.lib.uwo.ca/cjsotl_rcaceahttp://ir.li b.uwo.ca/cjsotl_rcacea/vol8/iss1/6

Thiagarajan, S., Semmel, D. S. & Semmel, M. I. (1974). Instructional Development for Training Teachers of Exceptopnal Children. *Indiana: Indiana Univer- sity Bloomington.* 192 Pages.

Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*. San Francisco: Jossey-Bass

UNESCO. (2017). *Education for Sustainable Development Goals: Learning Objectives*. Paris: UNESCO.

UNICEF. (2020). *Healthy Environments for Healthy Children: Global Programme Framework*. New York. https://www.unicef.org/media/91216/file/Healthy-Environments-for-Healthy-Children-Global-Programme-Framework-2021.pdf

United Nations. (2025). *Goal 6: Ensure access to water and sanitation for all*. United Nations Sustainable Development. Retrieved 1 March 2025, from https://www.un.org/sustainabledevelopment/water-and-sanitation/

Vigueras, G, A., & Sánchez-Trujillo, M. (2023). Study for the validation of an instrument for foreign language learning. *ECORFAN Journal Spain,* 10(19), 17-32. https://doi.org/10.35429/ejs.2023.19.10.17.32

Webb, F., Smith, C. & Worsfold, K. (2011). *Research skills toolkit.* http://www.griffith.edu.au/gihe/resources-suppport/graduate

Willison, J & O'Regan, K. (2007). Commonly known, commonly not known, totally unknown: A framework for students becoming researchers. *Higher Education Research and Development,* 26(4), 393–409.

Zainul, A. (2002). *Penilaian Hasil Belajar. Pusat Antar Universitas*, Direktorat Jenderal Pnedidikan Tinggi: Departemen Penddilan dan Kebudayaan.