Development of Critical Thinking Essay Test Instrument and Prosocial Intention Questionnaire for Environmental Care in Students

Muhammad Ikhsan[1,*], Sri Atun[2], Fauji Agusta[1], Habibatul Unayah[1], Rusdiman Buhera[1], Oky Pamungkas[1], Muhamad Sarip[3], Patar Ariyanto Sitorus[1]
[1]Department of Science Education, Universitas Negeri Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
[2]Department of Chemistry Education, Universitas Negeri Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
[3]Department of Biology Education, Universitas Negeri Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia

Article Info

ABSTRACT

Purpose of the study: Assessment of students' critical thinking skills and environmental care attitudes is crucial for science learning. This research aims to develop an essay test instrument for critical thinking and a questionnaire for environmental care attitudes, evaluating their validity through expert assessments and trials.

Methodology: Utilizing the Research and Development (R&D) model by Oriondo & Dallo-Antoio (1998), the study comprises three stages: test planning, trials, and validation. A random sampling technique was employed, involving 267 students. The critical thinking instrument consists of essay questions, while the environmental care attitudes are measured using a prosocial intentions questionnaire.

Main Findings: Validity results indicate a high Aiken index of 0.96 for critical thinking essays and 1.0 for the environmental care questionnaire. Reliability estimates were 0.73 for critical thinking and 0.71 for environmental attitudes, both good. The difficulty level for the critical thinking instrument was classified as medium (0.68), while the environmental care instrument was deemed difficult (1.65). Limitations include focusing solely on science subjects, trials restricted to one area, and reliance on self-reports for environmental attitudes, which may introduce bias.

Novelty/Originality of this study: The novelty of this study lies in integrating critical thinking assessment with environmental care attitudes, a rare approach in previous research. In conclusion, the developed instruments effectively measure students' critical thinking skills and environmental care attitudes in science education, particularly regarding global warming.

Corresponding Author:
Muhammad Ikhsan,
Department of Science Education, Universitas Negeri Yogyakarta,
Jl. Colombo No.1, Karang Malang, Caturtunggal, Sleman Regency, D. I. Yogyakarta, 55281, Indonesia
Email: elfitra@unimed.ac.id

1.   INTRODUCTION

Individuals must possess critical thinking skills to face the 21st century in order to compete globally. 21st-century learning includes skills that individuals must have, namely critical thinking, creativity, collaboration, communication, problem-solving, and digital literacy [1]. Science education plays a strategic role in developing students' critical thinking skills and environmental awareness in today's global era [2]. Science education provides students with the opportunity to construct their own concepts and offers direct experiences to explore and understand the surrounding environment scientifically [3], [4]. Therefore, students need to possess

critical thinking skills to achieve deep understanding and analysis in learning [5]-[7]. Additionally, science education is also related to fostering environmental awareness to observe the surrounding nature [8], [9].

Critical thinking involves individuals who conduct in-depth analysis, identify problems, evaluate information, and make rational decisions, as well as a reflective thinking process that focuses on deciding what to believe or do [10], [11]. Critical thinking skills in the learning process are essential, as students do not simply accept knowledge as it is, but rather go through a process of filtering that information so it can be accepted and understood [12]. Critical thinking is a cognitive and mental process that influences students to act in making decisions and solutions based on facts [13]. The critical thinking skills possessed by students consist of interpretation, analysis, explanation, evaluation, and conclusion [10], [13], [14]. Critical thinking can enhance students' understanding of learning materials because students are required to interpret, analyze, explain, and draw conclusions [15].

Environmental concern is a behavior or action that always strives to prevent damage to the surrounding natural environment and seeks to repair the damage that has already occurred [16]. Environmental concern must be possessed as demanded by the Sustainable Development Goals (SDGs) [17]. In the SDGs, there are 17 goals that countries must achieve for sustainable development, specifically point 13 on climate change and point 15 on terrestrial ecosystems. This means that an attitude of environmental concern is necessary to achieve effective climate change management and terrestrial ecosystem preservation [18]. The Ministry of Environment and Forestry mandates that students possess an attitude of environmental concern as stated in regulation number P.52/MENLHK/SETJEN/KUM1/9/2019 regarding the movement for environmental care and culture in schools, one of its programs being Adiwiyata (Green School), which aims to foster knowledge and awareness among school members in efforts to preserve the environment [19]. To achieve the status of an Adiwiyata-nominated school, it is essential to instill an attitude of environmental concern, not only to attain the nomination status but also for the cleanliness of the school environment [20].

Environmental awareness plays an important role in science education in understanding ecosystem concepts, introducing environmental issues such as global warming, pollution, habitat destruction, and developing awareness of biodiversity [20], [21]. The environmental awareness that students should possess and apply both at school and at home includes energy conservation, waste management, carbon emission reduction, and environmental care [19].

Global warming Global warming is one of the most complex and urgent environmental challenges of the 21st century that requires ways to address it from various aspects of life, including the world of education [22], [23] Global warming is the event of rising average temperatures in the atmosphere and on the Earth's surface, which has adverse effects on the Earth from various aspects if it occurs [24]. Global warming is one of the science subjects taught in seventh grade. Global warming is a complex issue that requires high-level critical thinking skills [25]-[27]. Students must cultivate an environmental care attitude and think critically about the impacts of global warming [28], [29]. In the research [30] global warming is highly relevant and suitable for measuring students' environmental concern..

Critical thinking and environmental awareness require measurement tools to assess the levels of categories possessed by students. This measurement serves to validate and ensure the feasibility of the tools used [31]. Having valid and good measurement tools is very important to ensure that the results obtained are trustworthy and accurately represent the abilities possessed by the students [32]. In this study, the development of instruments to assess students' critical thinking skills and environmental awareness is discussed. The role of assessment in measuring critical thinking skills is very important, and there are differing opinions on how to measure students' critical thinking abilities. According to Nurulhasni et al, an instrument in the form of an essay was developed to measure critical thinking skills [12]. According to Adhitya et al, an instrument in the form of essays and multiple-choice questions was developed to measure critical thinking skills. The instrument developed by Nurulhasni et al, is based on six indicators: interpreting, formulating problems, analyzing, concluding, evaluating, and developing strategies [12]. In this study, the instrument used is an essay to measure and assess students' critical thinking skills. The advantage of using essays to measure critical thinking skills is that they encourage students to develop in-depth analysis, synthesize information from various sources, and construct arguments logically and systematically [34]. The weakness of essays is that they require a long time to assess the answers [35]. The measurement of environmental concern attitudes shows differences between previous studies. In the study Hariyadi and Maryani, a questionnaire was used to measure students' environmental concern [36]. According to Efriyani, the instrument was developed using multiple-choice questions so that students could understand and choose the actions or attitudes they should take [37].

The instrument used in this study must meet certain criteria to be deemed suitable, namely validity, reliability, and difficulty level [38]. Alderman et al, states that the validity test includes content validity and construct validity tests [39]. The index proposed by Aiken can be used. The construct validity of the research uses IRT. (Item Respon Theory). Item analysis using the Quest IRT program based on the RM model (Rasch Model). The research instrument designed for critical thinking skills includes essays and environmental awareness attitudes using a prosocial intention questionnaire. The prosocial intention questionnaire measures a

person's tendency or intention to perform positive actions that benefit others or society voluntarily. The main focus of this instrument is to identify internal motivations and the potential of individuals to engage in constructive social actions [40].

Although previous research has developed various instruments to measure critical thinking skills and environmental care attitudes, there are several gaps that need to be addressed. First, there are still limited instruments that specifically integrate these two aspects in the context of science learning, especially on the topic of global warming. Second, there are not many instruments that use the prosocial intention scale approach to measure environmental care attitudes, which can provide a more comprehensive picture of students' behavioral tendencies.

The urgency of this research is driven by several critical factors. First, the increasing global environmental challenges that require young people with critical thinking skills and high environmental awareness. Second, the need for valid and reliable measurement instruments to evaluate the effectiveness of science learning in developing these two competencies. Third, the demands of implementing the Adiwiyata program which requires appropriate measurement tools to assess the development of students' environmental care attitudes. To solve these problems, this research developed an instrument that combines an essay test to measure critical thinking skills and a prosocial intention scale questionnaire to measure environmental care attitudes. This approach was chosen because of its ability to measure higher-order thinking skills and behavioral tendencies in more depth. The instrument was developed through a systematic process with attention to aspects of validity, reliability, and appropriate level of difficulty.

The main objectives of this study were: (1) to develop an essay test instrument to measure students' critical thinking ability in science learning on the topic of global warming, (2) to develop a prosocial intention scale questionnaire to measure students' environmental care attitude, (3) to test the validity, reliability, and difficulty level of the developed instrument, and (4) to produce an instrument that can be used effectively in the evaluation of science learning to improve students' critical thinking ability and environmental care attitude.

2. RESEARCH METHOD

The research method used is Research and Development. This research model is the Test Development Model. The final product of this research is an instrument to measure critical thinking skills and environmental care attitudes in learning science material on global warming. Instruments of critical thinking ability and environmental care attitude are implemented from the results of the Oriondo & Dallo-Antoio test development model (1998) which consists of three stages, namely planning the test, testing the test, determining validity and reliability.
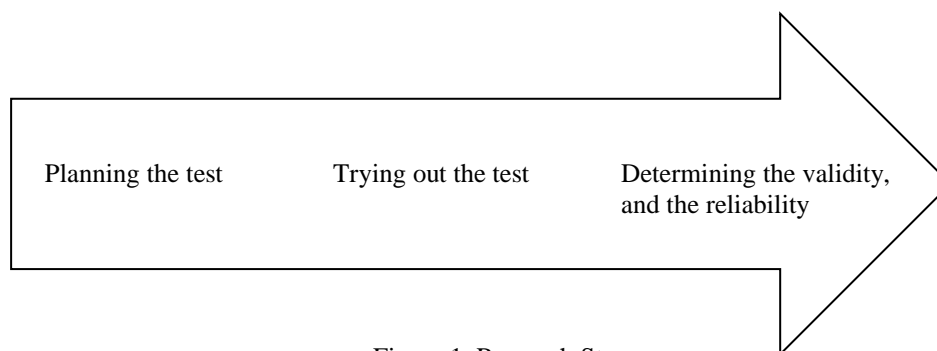

Figure 1. Research Steps

Planning a test involves determining the objectives, test materials, through test blueprints, writing test items and questionnaires, developing scoring guidelines, validating test items, and revising test items. The purpose of this is to measure students' critical thinking abilities and environmental care attitudes. The topic material chosen is global warming in the form of essay tests for critical thinking abilities and a prosocial intention questionnaire for environmental care attitudes. In this study, the scoring guidelines use an analytic scoring rubric and a Likert scale. Content validation was conducted by experts and analyzed using the Aiken index. The Aiken V formula is as follows:

$$V = \frac{\sum s}{(n(c-1))} \quad \dots (1)$$

Description:

V	: Raters agreement index
S	: The score given by each rater is reduced by the lowest score in the category used
(s : r - Io) r	: The score of the category chosen by the rater
Io	: The lowest score in the assessment category

n	: Total number of raters
c	: Total number of categories that are raters

The Aiken V index is a number that indicates the validity level of an item based on expert assessment, with a value range between 0 and 1. The categorization of Aiken V can be seen in Table 1 [41].

Table 1. Aiken V Category

Score of Aikens'V	Category
V > 0.8	High
$0.4 \leq V \leq 0.8$	Medium
V < 0.4	Low

The test trial was conducted with a research sample consisting of 267 junior high school students. The sampling technique used was random sampling. The sample size in the trial was based on Bond & Fox (2007), which states that the number of participants in IRT analysis starts from 30-300 participants.

Data analysis was conducted using the QUEST program which presents the reliability value of a test item. Able to evaluate instruments for critical thinking skills and environmental care attitudes, focusing on validity, reliability, and item difficulty. Item fit was assessed using Infit MNSQ, with criteria shown in table 2 [42].

Table 2. Item Requirements Suitable for the Rasch Model

Mean Square (MNSQ) Value	Information
>1.30	Not suitable for the Rasch model
0.77 – 1.30	Fits the Rasch model
<0,77	Not suitable for the Rasch model

The reliability of the Rasch Model using the QUEST program is reflected in the reliability of estimate item and reliability of cafe estimate. The criteria for reliability values are presented in Table 3 [12].

Table 3. Reliability Index Category Alpha Cronbach

Alpha Cronbach	Category
>0,9	Very High
0.70 – 0.90	High
0.40 – 0.70	Moderat
0.20 – 0.40	Low
<0,20	Very Low

The QUEST program also analyzes the difficulty level (threshold) of each item by reviewing the estimated item analysis results. (ambang batas). The analysis criteria range from -2.0 to 2.0, where items are classified as easy if their range or distribution is < -2.0, while items with a distribution range >2.0 fall into the difficult category. The criteria for determining the difficulty level of items are presented in Table 4 [42].

Table 4. Threshold Category

Threshold Value	Category
b > 2	Very Difficult
$1 < b \leq 2$	Difficult
$-1 < b \leq 1$	Moderate
$-1 < b \geq -2$	Easy
b < -2	Very Easy

3. RESULTS AND DICUSSION

Research that applies the 4-D development model goes through several stages. The objective is to describe the results of developing teaching materials that use statistics and their effectiveness in increasing students' statistical literacy and self-efficacy. The following are the results of the research based on the stages of the research model:

Planning the Test

Test planning involves determining the objectives, test materials, through test blueprints, writing test items and questionnaires, creating scoring guidelines, validating test items, and revising test items. The purpose

of this research is to measure students' critical thinking skills and environmental concern attitudes through essay instruments and prosocial intention scale questionnaires. The topic material used is global warming for junior high school students. The following are the indicators of critical thinking and environmental concern in tables 5 and 6.

Table 5. Critical Thinking of Indicator

Indicator	Sub Indikator	Items
Interpretation	Students are able to interpret the relationship and long-term impact on climate change between the increase in CO2 emissions and the change in global average temperature based on the data table	1
	Students are able to interpret the sectoral graphs of CO2 emission causes in 2010 and 2020 and compare them	2
Analysis	Students are able to analyze the positive and negative impacts on greenhouse gas emissions from the presented case.	3
	Students are able to analyze the main causes of global warming from the presented case	4
Evaluation	Students are able to provide evaluative solutions related to the issues in Yogyakarta with the city's hot temperature conditions	5
	students are able to explain and evaluate the policy of reducing usage Chloro Fluoro Carbon (CFC)	6
Explaning	Students are able to explain the impact of deforestation on the increase in global warming	7
	Students are able to explain how the use of alternative energy can address global warming	8
Conclusion	Students are able to draw conclusions regarding the steps to address global warming	9
	Students are able to draw conclusions about the social and economic impacts on coastal communities experiencing global warming.	10

Table 6. Environmental Care Attitude of Indicator

Indicator	Items
Energy Savings	1, 2
Waste Management	3,4,5
Carbon Emission Reduction	6,7
Environmental Maintenance	8,9,10

The instrument for environmental concern attitudes is in the form of a questionnaire using a prosocial intention scale, one example of the questionnaire items with the prosocial scale can be seen in Table 7.

Table 7. Prosociability Intensity Scale Questionnaire Items along with their Evaluation Scale

Item	Answer	
In a household, electrical appliances such as lights, televisions, and phone chargers often remain on even when not in use. This not only causes energy waste but also contributes to higher electricity bills and negative environmental impacts. Every individual can play a role in saving energy and reducing their carbon footprint, if this happens in your home, then what you should do is. . . .	A. Turn off the unused electrical appliances.	4 (Very Good)
	B. Telling the family to turn off the electrical appliances	3 (Good)
	C. Scolding the family for not being wasteful with electricity usage	2 (Less Good)
	D. Leaving the electrical appliances on	1 (Not Good)

Test Instrument

The instrument for testing critical thinking skills in the form of an essay with 10 items and environmental awareness in the form of a pro-social intention questionnaire with 10 items was then validated by experts. The results of the instrument validation using Aiken V are as follows in tables 8 and 9.

Table 8. Critical Thinking Validation Results

No	Validator	The average of the assessments
1	I	3,6
2	II	3.8
3	III	4,0
4	IV	4,0
5	V	4,0
Average		3,88
Aiken V		0,96
Category Validity		High

Based on table 7, all items have high results, with a total of 10 items there are 6 items with an Aiken V index of 0.93 and 4 items with an Aiken V index. Overall, the critical thinking inventory of items with an Aiken V index of 0.96 is in the high category based on table 1. There are also some revisions and input related to the use of words to be easily understood by students and the assessment rubric.

Table 9. Environmental Care Attitude Validation Results

No	Validator	The average of the assessments
1	I	4,0
2	II	3.9
3	III	4,0
4	IV	4,0
5	V	4,0
Average		3,98
Aiken V		1
Category Validity		High

Based on table 9, all questionnaire items have high results, with a total of 10 items there is 1 item with an Aiken V index of 0.93 and 9 items with an Aiken V index of 1. Overall, the critical thinking instrument items with an Aiken V index of 1 are in the high category based on table 1. There are several revisions and input, namely the use of the word fund to change the answer to the prosocial intention scale.

Test Result

The instrument developed was then tested on junior high school students with a sample size of 267. Then analyzed using the Quest program. The Quest program evaluates item fit using the Infit Mean Square (Infit MNSQ) mean and standard deviation. Items are considered to fit the model if their Infit MNSQ values are between 0.77 to 1.30 and Infit t values range from -2 to +2 [42]. The results of item fit and Infit Mean Square (Infit MNSQ) critical thinking in Figure 2.

```
Berpikir Kritis
-------------------------------------------------------------------------------
Item Fit                                                      23/11/24 21:57
all on all (N = 267 L = 7 Probability Level= .50)
-------------------------------------------------------------------------------
INFIT
 MNSQ         .56      .63      .71      .83     1.00     1.20     1.40     1.60     1.80
-----------------+---------+---------+---------+---------+---------+---------+---------+---------
   1 item 1                            .        *    |                  .
   2 item 2                            .           * |                  .
   3 item 3                            .             |            .   *
   4 item 4                            .           * |                  .
   5 item 5                            .             |       *          .
   6 item 6                   *        .             |                  .
   7 item 7                            .             |       *          .
=================================================================================
```
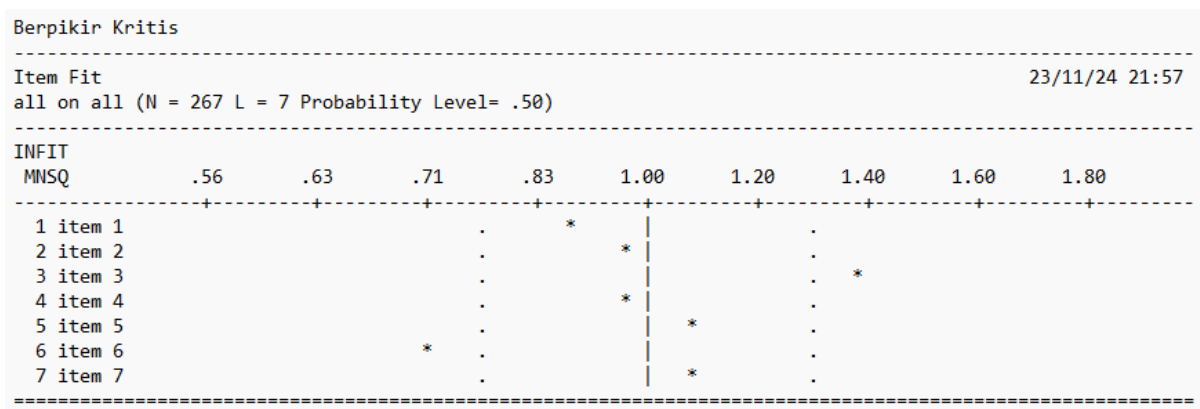
Figure 2. Results of Item Fit and Infit MNSQ Critical Thinking Essay

Figure 2, shows the results of item fit for each item. The items used in this instrument are essays. MNSQ infit value 0.77-1.33. Seen in Figure 2 items that are outside the provisions of the MNSQ infit value or invalid items in item 3 (1.37) interpretation indicator and item 6 (0.72). Item 1 (0.88) interpretation indicator, item 2 (0.97) analysis indicator, item 4 (0.97) evaluation indicator, item 5 (1.08) explanation indicator and item 7 (1.08) conclusion indicator are included in the validity criteria items. From the results of validation by experts and then analyzed using Aiken V, there were 10 items with high categories but researchers used 7 items for testing. Researchers consider the effectiveness of assessment, time management, quality of answers and consideration of student abilities. This is in line with research [43] using test instruments in the form of essays to measure students' cognitive is not effective with a large number of items, this can create student saturation in answering these items. According to research [44] revealed that the use of test instruments with a large number of items requires a long time estimation, in terms of students and their assessment. The number of 2 invalid items in item 3 and item 6 items with double indicators in item 1 and item 7, namely the interpretation indicator and the conclusion indicator. A total of 5 items are valid to be used to measure students' critical thinking skills. Furthermore, for the reliability results with the MNSQ infit value in table 10.

Table 10. Reliability and Infit MNSQ of Critical Thinking Essay

No	Reability	Value	Fit Statistics	Value
1	Reability of Estimate	0,73	Infit MNSQ	0,81-1,21
2	Reability of Case Estimate	0,71	Infit MNSQ	0,41-1,57

Table 10, shows the results of critical thinking reliability and MNSQ infit fit with the Rasch Model. Based on table 3, the reliability of estimate value is 0.73 in the high category and the reliability of case estimate is 0.71 in the high category. The higher the resbility of estimate value, the more accurate and reliable this instrument is to measure students' critical thinking, as well as the rebility of case estimate, the higher the value, the more accurate this instrument is to be measured in each student with different abilities [45]. Then the fit with the Rasch Model based on table 2 range provisions, the results of the MNSQ infit of estimate in the range of 0.81 - 1.21 are in accordance or fit with the Rash Model, which is good and accurate for use in measuring critical thinking. The results of the MNSQ infit of case estimate in the range 0.41-1.57 are not suitable or do not match the Rasch Model, this requires revision or disposal of items, there are some students who are some items that are not accurate, high ability students answer the question incorrectly and low ability students answer the question correctly. This is in line with research [46] that when the MNSQ infit results are outside the range of 0.77-1.33 there are some items that must be revised or discarded, it does not mean that all items cannot be used for measurement but there are some items that are not in accordance with the Rasch Model, to see which items are not suitable can be seen in the results of the fit item image or item estimate that is outside the range of 0.77-1.33. Furthermore, to see the distribution of the item estimate (threshold) in Figure 3 and the value of the item difficulty level in Table 11.
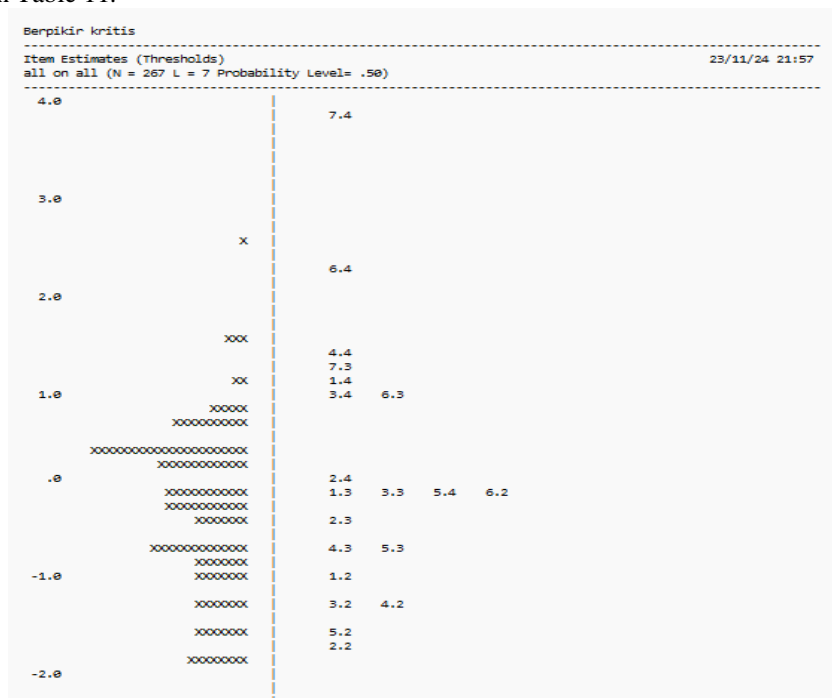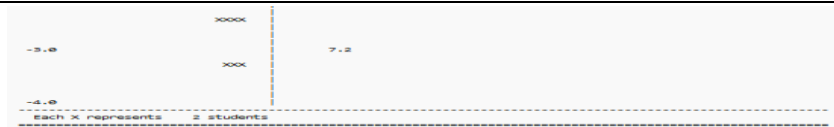
Figure 3. Critical Thinking Item Estimate (Tresholds)

Table 11. Level of Difficulty of Critical Thinking Essay

Items	Value of Level of Difficulty	Category
1	0,48	Medium
2	-0,38	Medium
4	1,26	Difficult
5	-0,01	Medium
7	2,03	Very Difficult
Average	0,68	Medium

Table 11, shows the results of the level of difficulty of items, as a whole essay test instrument to measure students' critical thinking with a value of 0.68 medium category based on the provisions of the threshold value in table 4. The category of very difficult items (2.03) in item 7 with the conclusion indicator, the category of difficult items (1.26) in item 4 with the evaluation indicator and the medium category in item 1 (0.48) interpretation indicator, item 2 (-0.38) analysis indicator and item 5 (-0.01) explanation indicator. Then for the results of item fit and Infit Mean Square (Infit MNSQ) environmental care attitude in Figure 4.
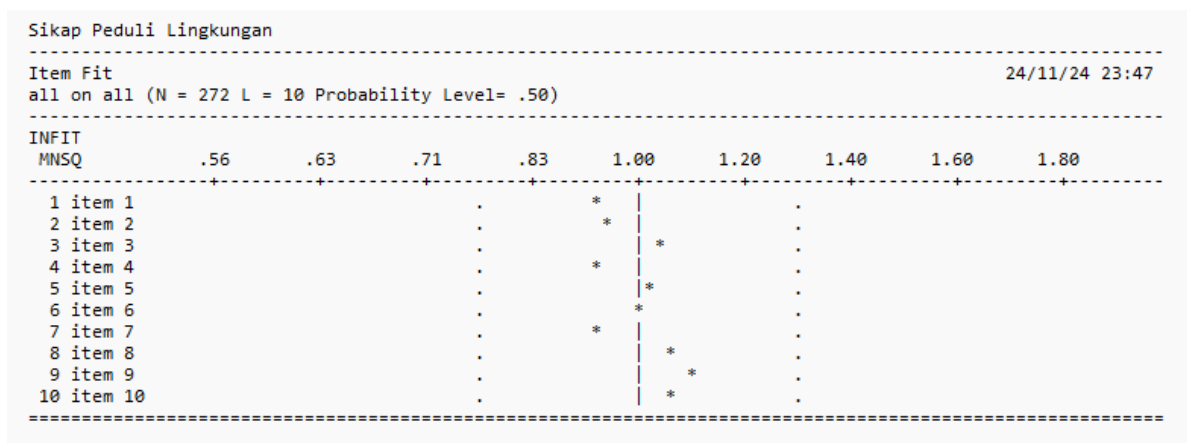

Figure 4. Results of Item Fit and Infit MNSQ Questionnaire for Environmental Care Attitude

Figure 3, shows the results of item fit for each item. The infit MNSQ value is 0.77-1.33. The item used in this item is a questionnaire. It can be seen in Figure 3 that all items are included in the provisions of the MNSQ infit value, which means that all items are valid or on the validity criteria. Item 1 (0.92) energy saving indicator, item 2 (0.94) energy saving indicator, item 3 (1.04) waste management indicator, item 4 (0.93) waste management indicator. Item 5 (1.03) waste management indicator, item 6 (1.00) carbon emission reduction indicator, item 7 (0.93) carbon emission reduction indicator, item 8 (1.05) environmental care indicator, item 9 (1.10) environmental care indicator and item 10 (1.06) environmental care indicator. From the results of validation by experts and analyzed using Aiken V, researchers used all items to be tested because the items tested for environmental care attitudes were in the form of questionnaires. Furthermore, for the reliability results with the MNSQ infit value in table 12.

Table 12. Reliability and Infit MNSQ of Questionnaire on Environmental Care Attitude

No	Reability	Value	Fit Statistics	Value
1	Reability of Estimate	0,82	Infit MNSQ	0,94 - 1,06
2	Reability of Case Estimate	0,71	Infit MNSQ	0,56 - 1,42

Table 12, shows the results of the reliability of environmental care attitudes and infit MNSQ fit with the Rasch Model. Based on table 3, the reliability of estimate value is 0.82 in the high category and the reliability of case estimate is 0.71 in the high category. The higher the resbility of estimate value, the more accurate and

reliable this instrument is to measure students' critical thinking, as well as the rebility of case estimate, the higher the value, the more accurate this instrument is to be measured in each student with different abilities [45]. Then the fit with the Rasch Model based on table 2 provisions range, the results of infit MNSQ of estimate in the range of 0.94 - 1.06, which is appropriate or suitable with the Rash Model, which is good and accurate for use in measuring environmental awareness. The results of infit MNSQ of case estimate in the range of 0.56-1.42 are not suitable or do not match the Rasch Model, this requires revision or disposal of items, there are some students who are some items that are not accurate, high ability students answer the question incorrectly and low ability students answer the question correctly. This is in line with research [46] that when the MNSQ infit results are outside the range of 0.77-1.33 there are some items that must be revised or discarded, it does not mean that all items cannot be used for measurement but there are some items that are not in accordance with the Rasch Model, to see which items are not suitable can be seen in the results of the fit item image or item estimate that is outside the range of 0.77-1.33. Furthermore, to see the distribution of the item estimate (threshold) in Figure 4 and the value of the results of the item difficulty level in Table 13.
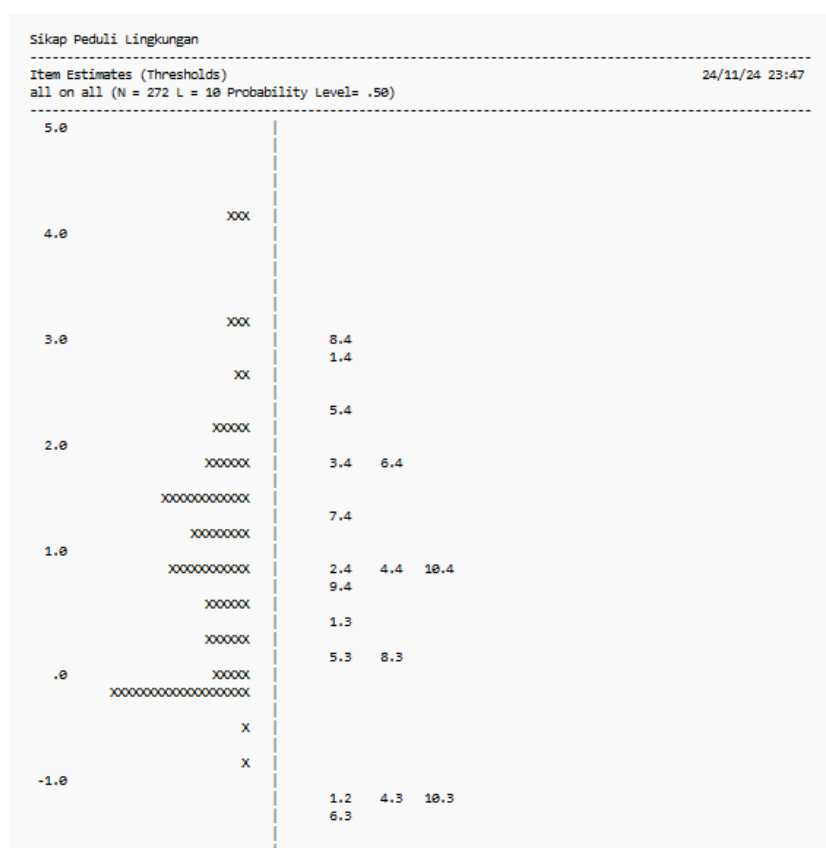

Figure 4. Item Estimate (Treshold)

Table 13. Level of Difficulty of Questionnaire on Environmental Care Attitude

Items	Value of Level of Difficulty	Category
1	1,83	Difficulty
2	1,38	Difficulty
3	1,77	Difficulty
4	0,79	Medium
5	2,44	Very Difficulty
6	1,43	Difficulty
7	1,87	Difficulty
8	2,71	Very Difficulty
9	1,54	Difficulty
10	0,76	Medium
Average	1,65	Difficulty

Table 13, shows the results of the level of difficulty of items, as a whole questionnaire instrument to measure students' environmental care attitudes with a value of 1.65 difficult categories based on the provisions of the threshold value in table 4. The category of items that are very difficult in item 8 (2.71) with environmental care indicators, item 5 (2.44) indicators of waste management. The category of difficult items in item 1 (1.83) energy saving indicator, item 2 (1.38) energy saving indicator, item 3 (1.77) waste management indicator, item 6 (1.43) carbon emission reduction indicator, item 7 (1.87) carbon emission reduction indicator, item 9 (1.54) environmental care indicator. Moderate category in item 4 (0.79) indicator of waste management and item 10 (0.76) indicator of environmental care.

Characteristics of essay instruments to measure critical thinking present contextual problems in everyday life about rising air temperatures, the environment and global warming. The characteristics of the questionnaire instrument to measure environmental care attitudes in the form of a prosocial intention scale, students are presented with answers to lead to action towards environmental conditions related to global warming. The development of critical thinking instruments in the form of essays was studied by [12], [34], [43] essay tests are considered to be able to measure complex cognitive levels and encourage students to activate higher-level thinking skills in answering questions. The development of a questionnaire environmental care attitude instrument was researched by [36], [37], [47].

The advantages of the critical thinking essay instrument and the prosocial intention scale questionnaire for environmental care attitudes have high reliability values and meet the validity of theory and empirical testing. The weaknesses of the critical thinking essay instrument are the objectivity of the assessment, scoring guidelines and the time needed to evaluate it which is long. The results of this study can be used to measure students' critical thinking skills and environmental care attitudes so that it can be used as an evaluation of science learning that equips students' abilities.

The results of this study indicate that the instruments developed to measure students' critical thinking skills and environmental care attitudes have high validity and reliability. The interpretation of the data results showed that out of a total of 10 validated items, 5 items were declared valid for use in measuring critical thinking skills. This is in line with the research findings [48], [49] which shows that essay-based instruments are effective in evaluating students' critical thinking skills, especially in the context of science learning related to environmental issues such as global warming [50]. Based on the results of the analysis, it can be seen that the use of essay-based instruments and prosocial intention questionnaires can effectively measure students' critical thinking skills and environmental care attitudes. This shows the importance of developing relevant measurement tools in science education to increase environmental awareness among students [51]. This research also confirms that an approach that combines theory and practice in instrument development can produce more valid and reliable measuring instruments.

The implication of the results of this study is the need for integration of critical thinking skills and environmental care attitudes in the science education curriculum. The development of valid and reliable instruments not only helps in the evaluation of students' abilities but also encourages more effective teaching in the context of environmental issues. Thus, teachers need to be trained to use these instruments optimally in the learning process. Limitations of this study include a sample limited to junior high school students in one location, so the results may not be generalizable to a wider population. In addition, the time required for essay grading may be an obstacle for teachers in a classroom situation with a large number of students.

Based on the results of the study, it is recommended that further research be conducted involving a more diverse sample from various locations to increase the generalizability of the findings. In addition, the development of additional instruments, such as online-based questionnaires or other measurement tools, can be considered to facilitate the assessment process. Finally, training for educators on the use of this instrument is essential to ensure effectiveness in classroom application.

4.   CONCLUSION

Critical thinking instrument in the form of essay and environmental care attitude in the form of a questionnaire with prosocial intention scale. Validation based on Aiken index 0.96 critical thinking and 1 environmental care attitude. Reliability value of estimate 0.73, reliability of case estimate 0.71 critical thinking and reliability value of estimate 0.82, reliability of case estimate 0.71 environmental care attitude. The level of difficulty of the critical thinking instrument is 0.68 in the medium category and the environmental care attitude instrument is 1.65 in the difficult category. The results of the development of the critical thinking essay instrument consisted of 5 items that met the validity and reliability of the 10 items compiled. The environmental care attitude questionnaire consists of 10 items that meet the validity and reliability.

REFERENCES

[1] S. Hanipah, "Analisis kurikulum merdeka belajar dalam memfasilitasi pembelajaran abad ke-21 pada siswa menengah atas [Analysis of the independent learning curriculum in facilitating 21st century learning for high school students]," *Jurnal Bintang Pendidikan Indonesia*, vol. 1, no. 2, pp. 264-275, 2023, doi: 10.55606/jubpi.v1i2.1860.

[2] L. Kartina, N. Nurmeimana, I. Abamba, and A. J. Aswad, "Ability analysis of providing explanations and strategies in learning science," *Journal Evaluation in Education (JEE)*, vol. 4, no. 3, pp. 88–94, 2023, doi: 10.37251/jee.v4i3.699.

[3] A. S. Wahyuni, "Literature review: Pendekatan berdiferensiasi dalam pembelajaran IPA [Literature review: Differentiated approach in science learning]," *Jurnal Pendidikan Mipa*, vol. 12, no. 2, pp. 118–126, 2022, doi: 10.37630/jpm.v12i2.562.

[4] A. Y. Septajati and A. Widowati, "Development of science electronic modules with an ethnoscience integrated STEM approach in traditional herbal medicine making," *Journal of Educational Science and Technology (EST)*, vol. 10, no. 2, pp. 161–170, 2024.

[5] M. A. Putri and P. A. L. Simbolon, "Model pembelajaran problem solving dan korelasinya terhadap kemampuan berpikir kritis siswa [Problem solving learning model and its correlation with students' critical thinking skills]," *Journal Evaluation in Education (JEE)*, vol. 3, no. 3, pp. 87–91, 2022, doi: 10.37251/jee.v3i3.263.

[6] A. T. Kartika, L. Eftiwin, M. F. Lubis, and A. Walid, "Profil kemampuan berpikir kritis siswa kelas viii smp pada mata pelajaran IPA [Profile of critical thinking skills of grade VIII junior high school students in science subjects]," *Jurnal Riset Teknologi dan Inovasi Pendidikan*, vol. 3, no. 1, pp. 1–10, 2020.

[7] R. Cui and P. Teo, "Thinking through talk: Using dialogue to develop students' critical thinking," *Teach Teach Educ*, vol. 125, p. 104068, 2023, doi: 10.1016/J.TATE.2023.104068.

[8] J. Kukkonen, S. Kärkkäinen, and T. Keinonen, "Examining the relationships between factors influencing environmental behaviour among university students," *Sustainability (Switzerland)*, vol. 10, no. 11, 2019, doi: 10.3390/su10114294.

[9] Y. Li, D. Yang, and S. Liu, "The impact of environmental education at Chinese Universities on college students' environmental attitudes," *PLoS One*, vol. 19, no. 2 February, Feb. 2024, doi: 10.1371/journal.pone.0299231.

[10] Ennis R, "The Nature of Critical Thinking," 2011.

[11] R. Hidayah and M. Salimi, "Critical thinking skill: Konsep dan inidikator penilaian [Critical thinking skills: Concepts and assessment indicators]," *Taman Cendekia: Jurnal Pendidikan Ke-SD-an*, vol. 1, no. 2, pp. 127–133, 2020.

[12] D. Nurulhasni, E. Nurbaiti, R. S. Nabila, and J. Jumadi, "Developing and implementing an instrument for assessing critical thinking and visual representations in learning Physics materials of Optical Instruments," *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 1, no. 27, 2023, doi: 10.21831/pep.v27i1.53550

[13] F. Facione, "The Nature of Critical Thinking," 2011.

[14] P. A. Facione, *Permission to Reprint for Non-Commercial Uses Critical Thinking: What It Is and Why It Counts*. Peter A. Facione, Measured Reasons LLC, 2011.

[15] I. W. Distrik, C. Ertikanto, Y. S. Purwati, A. Saregar, and N. F. A. Rahman, "Digital Problem-Based worksheet with 3d pageflip: an effort to address concept understanding problems and enhance digital literacy skills," *Jurnal Pendidikan IPA Indonesia,* vol. 13, no. 1, pp. 116–127, 2024, doi: 10.15294/jpii.v13i1.48604.

[16] S. Suciati, "Environmental Conscious Attitudes of Geography Education Students," *Journal Evaluation in Education (JEE)*, vol. 5, no. 1, pp. 17–24, 2024, doi: 10.37251/jee.v5i1.841.

[17] M. Laras and H. Komalasari, "Meningkatkan sikap peduli lingkungan dengan implementasi sustainable development goals (SDGs) pada pembelajaran Tari Buyung [Improving environmental awareness by implementing sustainable development goals (SDGs) in Buyung Dance learning]," *Ringkang: Kajian Seni Tari dan Pendidikan Seni Tari*, vol. 4, no. 1, pp. 149-160, 2024.

[18] United Cities and Local Goverments, "Tujuan Pembangunan Berkelanjutan SDGs," 2019.

[19] Kementrian Lingkungan Hidup, "Peraturan Gerakan Peduli dan Berbudaya Lingkungan Hidup di Sekolah," 2022.

[20] M. Jen Ismail, "Pendidikan karakter peduli lingkungan dan menjaga kebersihan di sekolah [Character education on caring for the environment and maintaining cleanliness at school]," *Jurnal Pendidikan dan Pembelajaran* , vol. 4, no. 1, 2021.

[21] Z. Huang, Z. Gou, and S. Cai, "Energy justice in education sector: The impact of student demographics on elementary and secondary school energy consumption," *Heliyon*, vol. 9, no. 5, p. e16191, 2023, doi: 10.1016/j.heliyon.2023.e16191.

[22] J. Wang *et al.*, "Changing Lengths of the Four Seasons by Global Warming," *Geophys Res Lett*, vol. 48, no. 6, 2021, doi: 10.1029/2020gl091753.

[23] S. Alfonso, M. Gesto, and B. Sadoul, "Temperature increase and its effects on fish stress physiology in the context of global warming," *J Fish Biol*, vol. 98, no. 6, pp. 1496–1508, 2021, doi: 10.1111/jfb.14599.

[24] F. Mutlu and O. Nacaroğlu, "Examination of perceptions of gifted students about climate change and global warming," *Journal of Baltic Science Education*, vol. 18, no. 5, pp. 780–792, 2019, doi: 10.33225/jbse/19.18.780.

[25] R. H. Ristanto, A. S. Ahmad, and R. Komala, "Critical thinking skills of environmental changes: A biological instruction using guided discovery learning-argument mapping (GDL-AM)," *Participatory Educational Research*, vol. 9, no. 1, pp. 173–191, Jan. 2022, doi: 10.17275/per.22.10.9.1.

[26] G. Hillary, E. Djulia, and R. H. Hasibuan, "Analysis Critical Thinking Ability and Environmental Care Attitude of Junior High School Students on Global Warming Material," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 5, pp. 2383–2390, 2023, doi: 10.29303/jppipa.v9i5.2603.

[27] S. Chattuchai, A. Singseewo, and P. Suksringarm, "Development of knowledge, awareness, global warming decreasing behavior and critical thinking of grade 11 students using the four noble truths method with meta-cognitive techniques," *Educational Research and Reviews*, vol. 10, no. 16, pp. 2268–2275, 2020.

[28] O. Çavdar, "Investigation of secondary school students' knowledge levels about global warming in terms of various variables," *Educational Policy Analysis and Strategic Research*, vol. 18, no. 3, pp. 336–358, 2023, doi: 10.29329/epasr.2023.600.16.

[29] M. Asi, A. Retnoningsih, and A. Irsadi, "Effectiveness of interactive e-book global warming and climate change integrated socio scientific issues peat ecosystem," *Jurnal Penelitian Pendidikan IPA*, vol. 7, no. SpecialIssue, pp. 240–244, 2021, doi: 10.29303/jppipa.v7ispecialissue.1039.

[30] E. Suryawati, F. Suzanti, Zulfarina, A. R. Putriana, and L. Febrianti, "The implementation of local environmental problem-based learning student worksheets to strengthen environmental literacy," *Jurnal Pendidikan IPA Indonesia*, vol. 9, no. 2, pp. 169–178, 2020, doi: 10.15294/JPII.V9I2.22892.

[31] J. Wu, D. Long, N. Hafez, J. Maloney, Y. Lim, and H. Samji, "Development and validation of a youth climate anxiety scale for the Youth Development Instrument survey," *Int J Ment Health Nurs*, vol. 32, no. 6, pp. 1473–1483, 2023, doi: 10.1111/inm.13201.

[32] Y. Tanjung, T. Wulandari, F. Festiyed, Y. Yerimadesi, and Y. Ahda, "Development analysis of creative thinking test instruments on natural science materials," *Jurnal Pendidikan Fisika*, vol. 12, no. 1, p. 22, 2023.

[33] A. Adhitya, I. G. Astawan, and I. N. R. Adi, "Pengembangan instrumen penilaian keterampilan berpikir kritis dan berpikir kreatif berbasis google form [Development of critical thinking and creative thinking skills assessment instruments based on Google Forms]," *Jurnal Ilmiah Pendidikan Citra Bakti*, vol. 9, no. 2, pp. 250–261, 2022, doi: 10.38048/jipcb.v9i2.682.

[34] S. M. Danczak, C. D. Thompson, and T. L. Overton, "Development and validation of an instrument to measure undergraduate chemistry students' critical thinking skills," *Chemistry Education Research and Practice*, vol. 21, no. 1, pp. 62–78, 2020, doi: 10.1039/c8rp00130h.

[35] S. Qomariyah, R. Darmayanti, U. Rosyidah, and I. Ayuwanti, "Indicators and Essay Problem Grids on Three-Dimensional Material: Development of Instruments for Measuring High School Students' Mathematical Problem-Solving Ability," *JEMS: Jurnal Edukasi Matematika dan Sains*, vol. 11, no. 1, pp. 261–274, 2023, doi: 10.25273/jems.v11i1.14708.

[36] E. Hariyadi and E. Maryani, "Analisis instrumen literasi lingkungan siswa [Analysis of students' environmental literacy instruments]," *Gulawentah: Jurnal Studi Sosial*, vol. 6, 2021.

[37] Y. Efriyani, "Pengembangan instrumen penilaian sikap peduli lingkungan siswa sekolah dasar [Development of an assessment instrument for elementary school students' environmental care attitudes]," *Wiyata Dharma: Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 10, no. 2, pp. 157–166, 2022, doi: 10.30738/wd.v10i2.11177.

[38] W. Winarto, S. S., E. Cahyono, and W. Sumarni, "Developing a problem-solving essay test instrument (PSETI) in the instruction of basic science concepts in ethnoscience context," *Journal of Turkish Science Education*, vol. 19, no. 1, pp. 37–51, 2022, doi: 10.36681/tused.2022.108.

[39] H. Alderman, J. Friedman, P. Ganga, M. Kak, and M. Rubio-Codina, "Assessing the performance of the Caregiver Reported Early Development Instruments (CREDI) in rural India," *Ann N Y Acad Sci*, vol. 1492, no. 1, pp. 58–72, 2021, doi: 10.1111/NYAS.14543.

[40] S. Ampuni and S. B. S. Buwono, "Pengembangan dan identifikasi properti psikometris kuesioner intensi prososial berbasis skenario untuk anak dan remaja [Development and identification of psychometric properties of a scenario-based prosocial intention questionnaire for children and adolescents]," *Jurnal Psikologi Sosial*, vol. 20, no. 2, pp. 149–167, 2022, doi: 10.7454/JPS.2022.18.

[41] H. Retnawati, *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa, dan psikometrian) [Quantitative analysis of research instruments (guide for researchers, students, and psychometricians).]*. 2016.

[42] W. F. Hanna and H. Retnawati, "Analisis kualitas butir soal matematika menggunakan model rasch dengan bantuan software quest [Analysis of the quality of mathematics problem items using the Rasch model with the help of Quest software]," *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, vol. 11, no. 4, p. 3695, 2022, doi: 10.24127/ajpm.v11i4.5908.

[43] F. Larasati and S Syamsurizal, "Validitas instrumen tes keterampilan berpikir kritis peserta didik kelas XII SMA/MA tentang materi mutase [Validity of critical thinking skills test instrument for grade XII SMA/MA students on mutation material]," *Journal on Teacher Education*, vol. 4, pp. 365–372, 2022.

[44] Mayasari E, "Instrumen tes sebagai alat evaluasi [Test instruments as evaluation tools]," *Jurnal Pendidikan dan Pengajaran*, vol. 2, p. 55, 2023.

[45] A. Saputra, *CAMI: Aplikasi Uji Validitas dan Reliabilitas Instrumen Penelitian Berbasis Web CAMI: Web-Based Research Instrument Validity and Reliability Testing Application]*. Yayasan Ahmar Cendekia Indonesia, 2020.

[46] H. Putri, D. Susiani, and NS Wandani, "Instrumen penilaian hasil pembelajaran kognitif pada tes uraian dan tes objektif [Cognitive learning outcome assessment instruments in descriptive tests and objective tests]," *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, vol. 4, no. 2, 2022.

[47] Medina and Joneil B, "Description of environmental care analysis of students in elementary schools," *Journal Evaluation in Education (JEE)*, vol. 4, no. 3, pp. 95–103, 2023, doi: 10.37251/jee.v4i3.335.

[48] M. A. Putri and P. A. L. Simbolon, "Model pembelajaran problem solving dan korelasinya terhadap kemampuan berpikir kritis siswa [Problem solving learning model and its correlation with students' critical thinking skills]," *Journal Evaluation in Education (JEE)*, vol. 3, no. 3, pp. 87–91, 2022, doi: 10.37251/jee.v3i3.263.

[49] A. Doelvia, V. Thi, T. Hien, S. Rathee, A. Kebidanan, and A. Muara Bungo, "Assessment: The effectiveness of video media through the tiktok application on teenagers' knowledge about clean and healthy living behavior at junior high school level," *Journal Evaluation in Education (JEE)*, vol. 4, no. 4, pp. 168–174, 2023, doi: 10.37251/jee.v4i4.948.

[50] M. Maria, S. Silalahi, Y. Aggarwal, and U. Galadima, "Implementation of science student work sheet based on multiple intelligence materials temperature and their changes," *Journal Evaluation in Education (JEE)*, vol. 4, no. 3, pp. 104–109, 2023, doi: 10.37251/jee.v4i3.698.

[51] S. Mundarti and F. T. Aldila, "Affective assessment instrument based on krathwohl-anderson taxonomy in senior high school," *Journal Evaluation in Education (JEE)*, vol. 4, no. 2, pp. 74–79, 2023, doi: 10.37251/jee.v4i2.323.