JPPIPA 11(11) (2025)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Development of an Instrument to Assess Critical and Creative Thinking Skills in Fourth-Grade Elementary Students

Jufri Saputra[1*], Undang Rosidin[1], Dwi Yulianti[1], Mulyanto Widodo[1]

[1]Master Program in Primary School Teacher Education, University of Lampung, Bandar Lampung, Indonesia.

Received: July 3, 2025
Revised: September 12, 2025
Accepted: November 25, 2025
Published: November 30, 2025

Corresponding Author:
Jufri Saputra
JufriSaputra07@gmail.com

DOI: 10.29303/jppipa.v11i11.11994

Abstract: This study aims to develop a higher-order thinking skills-based assessment instrument to measure fourth-grade elementary students' critical and creative thinking in Science and Social Studies. Using the ADDIE R&D model, the instrument was validated by content, evaluation, and language experts, tested for readability by students, and evaluated for practicality by teachers. Results showed high validity (82%), very good readability (84.58%), and high practicality (93.6%). Student testing indicated stronger performance in critical thinking (interpretation, analysis, evaluation) than in creative thinking (flexibility, fluency, elaboration, originality). The instrument is thus valid, practical, and effective for identifying students' higher-order thinking skills.

Keywords: Assessment Instrument; Critical Thinking; Creative Thinking; Elementary School Students; Higher Order Thinking Skills

Introduction

The *Kurikulum Merdeka* emphasizes the development of 21st-century competencies, particularly critical thinking, creativity, communication, and collaboration (4C). These competencies are highly relevant in responding to global challenges that demand learners to think deeply, flexibly, and innovatively. At the elementary level, these skills are integrated into the subject of Natural and Social Sciences (*Ilmu Pengetahuan Alam dan Sosial*, or IPAS), which is designed to foster curiosity, active participation, and higher-order thinking skills (HOTS) among students.

Theoretically, critical and creative thinking are essential cognitive abilities that must be nurtured from an early age. Brookhart (2010) defines critical thinking as the ability to analyze, evaluate, and make reasoned decisions, while creative thinking involves flexibility, fluency, elaboration, and originality in generating ideas or solutions. Anderson and Krathwohl's (2001) revision of Bloom's Taxonomy places higher-order thinking skills—such as analyzing, evaluating, and creating—at the top levels of cognitive complexity, underscoring the importance of assessing these skills in learning processes.

Several previous studies support the significance of developing and measuring these skills. Yulianti dan Herpratiwi (2024), for instance, developed a SETS-based IPAS module that effectively enhanced students' critical thinking abilities. Similarly, Azharini et al. (2023) designed a HOTS-based assessment tool aimed at measuring students' critical and creative thinking within thematic learning.

However, a needs analysis conducted at SD Negeri 01 Bakung Udik revealed a gap between teachers' understanding of HOTS-based assessment and its classroom implementation. Although all teachers (100%) reported understanding the concept of HOTS, only 33.3% applied it in classroom practice, and merely 16.7% explicitly assessed students' critical and creative thinking abilities.

This situation indicates an urgent need for an assessment instrument that is valid, practical, readable, and aligned with the cognitive developmental stage of elementary school students. Therefore, this study aims

How to Cite:
Saputra, J., Rosidin, U., Yulianti, D., & Widodo, M. (2025). Development of an Instrument to Assess Critical and Creative Thinking Skills in Fourth-Grade Elementary Students . *Jurnal Penelitian Pendidikan IPA*, *11*(11), 909–918. https://doi.org/10.29303/jppipa.v11i11.11994

to develop an integrated and contextual HOTS-based assessment instrument to measure the critical and creative thinking skills of fourth-grade students in IPAS learning, and to support the effective implementation of the *Kurikulum Merdeka* at the primary education level.

Method

This study is a Research and Development (R&D) project that adopts the ADDIE model (Analysis, Design, Development, Implementation, Evaluation) with adjustments tailored to the development needs of a Higher Order Thinking Skills (HOTS) assessment instrument. The primary objective of this research is to develop an instrument capable of measuring elementary school students' critical and creative thinking skills. The study was conducted on January 21, 2025, at SD Negeri 01 Bakung Udik, located in Tulang Bawang Regency, Lampung Province, Indonesia. The target of this study was fourth-grade elementary school students, with research subjects comprising 31 fourth grade students, three fourth grade teachers, and six expert validators consisting of content experts, assessment experts, and language experts**.**

The research procedure followed five stages based on the ADDIE model: (1) the planning stage, which involved constructing the test blueprint, writing HOTS items, and developing scoring rubrics; (2) the validation and revision stage, in which the instrument was reviewed by experts and revised based on their feedback; (3) the limited trial stage, which tested the readability and practicality of the instrument through student and teacher responses; (4) the field trial stage, where the instrument was implemented with students to obtain empirical data; and (5) the product evaluation stage, which involved analyzing the instrument's validity, reliability, discrimination index, difficulty level, as well as its readability and practicality.

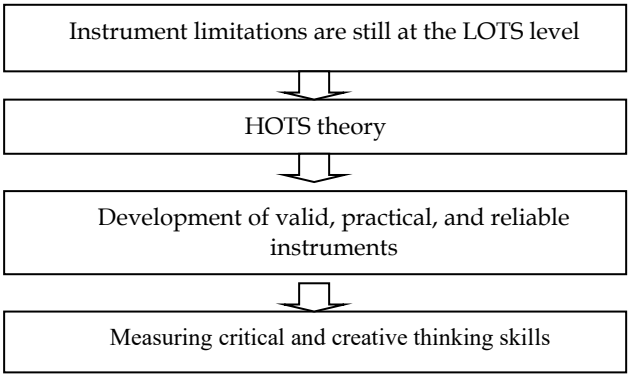


Figure 1. Research Flow Chart

The primary instrument used in this study is an essay-type test developed to measure students' critical and creative thinking abilities. The indicators of critical thinking in this study refer to the framework proposed by Karim & Normaya (2015), which is based on Facione's theory, as follows:

Table 1. Indicators of Critical Thinking Skills

Critical Thinking Indicator	Description
Interpretation	Understanding the problem demonstrated by accurately stating the given and asked information
Analysis	Identifying relationships between pieces of information and providing appropriate explanations
Evaluation	Applying appropriate strategies to solve the problem thoroughly and correctly
Inference	Drawing conclusions based on the available information

The indicators of creative thinking skills refer to the framework proposed by Munandar (2017), which includes:

Table 2. Indicators of Creative Thinking Skills

Creative Thinking Indicator	Description
Fluency	The ability to generate a large number of relevant ideas or responses to a given problem
Flexibility	The ability to generate various alternative approaches to solving a problem
Originality	The ability to provide unique or uncommon responses
Elaboration	The ability to elaborate or expand on an idea in a detailed manner

The main instrument used in this study was an essay-type test designed to assess critical thinking skills measured through indicators such as interpretation, analysis, evaluation, and inference and creative thinking skills measured through indicators such as fluency, flexibility, originality, and elaboration. Supporting instruments included expert validation sheets, student readability questionnaires, and teacher practicality questionnaires. Data were collected through two main approaches: tests and non-test techniques. The test was used to obtain quantitative data from students' answers, while the non-test techniques, such as questionnaires and validation sheets, were used to collect qualitative data from experts, teachers, and students regarding the validity, readability, and practicality of the instrument.

The collected data were analyzed using several techniques: expert validity was measured using Aiken's V formula; empirical validity was analyzed using

Pearson Product Moment correlation; reliability was tested using the Cronbach's Alpha formula. Additionally, item difficulty and discrimination indices were calculated to assess the quality of each test item. The practicality and readability of the instrument were analyzed using percentage-based analysis of the questionnaire responses from teachers and students.

Result and Discussion

This study aimed to develop a Higher Order Thinking Skills (HOTS) assessment instrument to measure the critical and creative thinking abilities of fourth-grade elementary school students. The findings are presented based on four key phases of the development process: expert validation, empirical testing, practicality, readability, and measurement results.

Expert Validation

Based on the results of each aspect of expert validation, it can be concluded that the instructional instrument meets the validity criteria. The recapitulation of validation scores from the subject matter expert, evaluation expert, and language expert is presented in Table 3.

Table 3. Expert Validation

Item	r1	r2	S1	S2	∑S	V
Evaluation Expert	24	26	18	20	10	0.79
Language Expert	25	29	19	23	42	0.85
Content Expert	25	26	19	20	39	0.81
Average						0.82
Category					Valid, complete, and usable	

Empirical Testing

The item validity test was conducted to determine whether each test item used in this study is valid or not. The validation was carried out on 12 respondents using 10 test items. The validity test in this study employed the Product Moment formula with the assistance of Microsoft Office Excel 2013. The results of the item validity test used in the study are presented in Table 4.

Table 4. Empirical Testing

Item Number	r Calculated	r Table	Description
1	0.759	0.576	Valid
2	0.764	0.576	Valid
3	0.701	0.576	Valid
4	0.680	0.576	Valid
5	0.545	0.576	Not Valid
6	0.742	0.576	Valid
7	0.623	0.576	Valid
8	0.771	0.576	Valid
9	0.760	0.576	Valid
10	0.117	0.576	Not Valid

Based on the results of the validity test conducted on 10 essay items, it was found that 8 items were declared valid, while 2 items were not valid. The two invalid items were removed from the instrument because they did not meet the predetermined validity criteria, namely $rcalculated \leq rtable$.

Reliability Testing of the Items

Reliability testing was conducted to determine the accuracy or consistency level of the items in this research instrument. The test was carried out on 12 respondents using 8 valid items and employed the Cronbach's Alpha method, with the assistance of Microsoft Office Excel 2013. The results of the reliability testing are presented in Table 5.

Table 5. Results of Item Reliability Testing

Reliability	N (Items)	Category
0.865	8	Very Strong

Based on Table 5, the results of the item reliability testing show a score of 0.865, which falls into the "very strong" category. This indicates that the test items possess high reliability and are suitable for use in the research.

Item Difficulty Testing

The item difficulty testing in this study aimed to determine whether each test item falls into the easy, moderate, or difficult category. The test was conducted with 12 respondents using 8 items. The analysis of item difficulty was carried out with the assistance of Microsoft Office Excel 2013. The results of the item difficulty analysis are presented in Table 6.

Table 6. Results of Item Difficulty Testing

Item Number	Difficulty Index	Category
1	0.316	Moderate
2	0.316	Moderate
3	0.300	Moderate
4	0.270	Difficult
5	0.316	Moderate
6	0.300	Moderate
7	0.300	Moderate
8	0.241	Difficult

Based on Table 6, the item difficulty analysis shows that the test items are categorized into two groups: difficult and moderate. Items number 4 and 9 fall into the difficult category, while items number 1, 2, 3, 6, 7, and 8 are classified as moderate. These results indicate that out of the total items analyzed, 2 items are considered difficult and 6 items are considered to have a moderate level of difficulty.

Item Discrimination Index Testing

The item discrimination index is an indicator used to categorize students into two groups: the upper group (students with high ability) and the lower group (students with low ability). This discrimination index testing was conducted with the assistance of Microsoft Office Excel 2013. The results of the item discrimination analysis are presented in Table 7.

Table 7. Results of Item Discrimination Index Testing

Item Number	Discrimination Index	Category
1	0.2	Moderate
2	0.2	Moderate
3	0.4	Moderate
4	0.4	Moderate
5	0.8	Very Good
6	0.2	Moderate
7	0.2	Moderate
8	0.8	Very Good

Based on Table 7, it is known that 6 items fall into the "Moderate" category and 2 items fall into the "Very Good" category.

Teacher Response Practicality Test

The practicality test involving three teachers at SD Negeri 01 Bakung Ilir aimed to determine the extent to which the HOTS instrument for the IPAS subject is practical for measuring students' critical and creative thinking skills. The assessment was conducted by three teachers and focused on three main aspects: (1) Attractiveness; (2) Ease of use; (3) Usefulness.

These three aspects were evaluated based on 10 statements. The teachers' responses to the HOTS instrument were then analyzed using descriptive percentage analysis, resulting in an average score of 93.6%. According to the practicality assessment criteria table, this score falls into the "Very Practical" category. The results of the teacher response practicality test are presented in Table 8.

Table 8. Results of the Teacher Response Practicality

Evaluated Aspect	Percentage	Category
Attractiveness	100%	Very Practical
Ease of Use	90%	Very Practical
Usefulness	91%	Very Practical
	93.6%	Very Practical

Student Response Readability Test

The readability test of student responses was conducted to determine the level of readability of the HOTS instrument product in the IPAS subject, aimed at measuring critical and creative thinking skills of elementary school students.

This test was administered to 10 fourth-grade students at SD Negeri 01 Bakung Udik. The evaluation was based on three indicators: (1) Presentation; (2) Language; (3) Content.

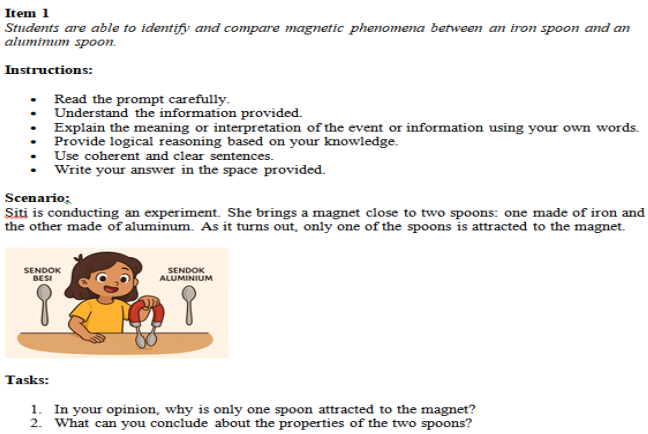
These three indicators consisted of 8 statements. The students' assessments of the HOTS instrument were then analyzed using descriptive percentage analysis, resulting in an average score of 84.58%. According to the readability assessment criteria table, the instrument falls into the "Very Good" category. The complete results of the student response readability test are presented in Table 9.

Table 9. Results of the Student Response Readability

Evaluated Aspect	Percentage	Category
Presentation	92.08%	Very Good
Language	81.67%	Very Good
Content	80.00%	Good
	84.58%	Very Good

Results of Critical Thinking Skills Assessment

The critical thinking instrument comprised four indicators, each measured through one test item:


Item 1
Students are able to identify and compare magnetic phenomena between an iron spoon and an aluminum spoon.

Instructions:

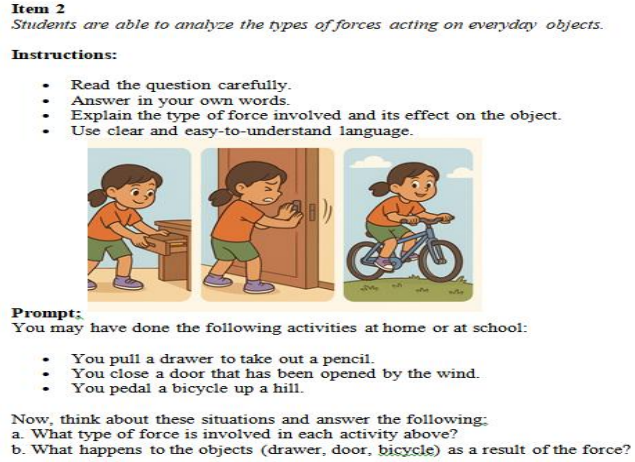
- Read the prompt carefully.
- Understand the information provided.
- Explain the meaning or interpretation of the event or information using your own words.
- Provide logical reasoning based on your knowledge.
- Use coherent and clear sentences.
- Write your answer in the space provided.

Scenario:
Siti is conducting an experiment. She brings a magnet close to two spoons: one made of iron and the other made of aluminum. As it turns out, only one of the spoons is attracted to the magnet.

Tasks:

1. In your opinion, why is only one spoon attracted to the magnet?
2. What can you conclude about the properties of the two spoons?

Figure 1. Interpretation Indicator


Item 2
Students are able to analyze the types of forces acting on everyday objects.

Instructions:

- Read the question carefully.
- Answer in your own words.
- Explain the type of force involved and its effect on the object.
- Use clear and easy-to-understand language.

Prompt:
You may have done the following activities at home or at school:

- You pull a drawer to take out a pencil.
- You close a door that has been opened by the wind.
- You pedal a bicycle up a hill.

Now, think about these situations and answer the following:
a. What type of force is involved in each activity above?
b. What happens to the objects (drawer, door, bicycle) as a result of the force?

Figure 2. Analysis Indicator

Item 3
Students are able to evaluate the effects of force on the motion or shape of an object.

Instructions:

- Read the story about Siti and Dani's experiments.
- Choose the experiment that, in your opinion, best demonstrates the effect of force on an object.
- Write down whose experiment you chose (Siti or Dani).
- Explain why you chose that experiment. Provide a clear reason.
- Write your answer in your own words and in an organized manner.
- There is no wrong answer—as long as you can explain your reasoning clearly.

Scenario:
Siti and Dani conducted two different experiments.

- Siti pressed a rubber ball. The ball changed shape and then returned to its original form.
- Dani pushed a toy box. The box moved and then stopped.

In your opinion, whose experiment best shows the effect of force?
State your choice, and explain your reason in your own words.

Figure 3. Evaluation Indicator

Item 4
Students are able to analyze types of landforms and their relation to specific professions.

Instructions:

- Read the question carefully.
- Observe whether there are rice fields or rice farmers in your area.
- Think about why farmers in lowland areas are more likely to grow rice than other crops.
- Write your answer in complete and clear sentences.
- Use your own experiences and knowledge about your local environment.
- Explain your reasoning in your own words.

Question:

Are there rice fields or rice farmers in your area? In your opinion, why do farmers in lowland areas more often grow rice than other crops? Explain your reasoning clearly.

Figure 4. Interpretation Indicator

The results of the critical thinking assessment show varied performance across the four indicators. For the Interpretation indicator (Item No. 1), a total of 11 students scored 3 and 4 students scored 4, indicating that the majority demonstrated medium to high levels of interpretation ability. In the Analysis indicator (Item No. 2), the most frequent scores fell into the medium category (score 3, 16 students) and low category (score 2, 9 students), suggesting that most students were able to analyze information adequately, though not optimally. For the Evaluation indicator (Item No. 3), the majority of students scored 2 (13 students) and 1 (9 students), indicating that their evaluative skills were still relatively low. Lastly, the Advanced Interpretation indicator (Item No. 4) revealed that most students (15 students) scored only 1, suggesting that the item was quite difficult and that students faced challenges in interpreting information at a deeper level.

Results of Creative Thinking Skills Assessment

The creative thinking instrument also comprised four indicators, each measured through one test item:

Item 5
Students are able to explain magnetic and non-magnetic phenomena in relation to magnets across various situations.

Instructions:

- Read the question carefully.
- Identify the objects mentioned in the question.
- Consider the properties of magnets and the types of materials that can be attracted to them.
- Write down which objects are attracted to magnets and explain your reasoning.
- Provide one additional example of an object around you that is also attracted to magnets.

Question:

Budi found several objects at home, namely a metal spoon, an eraser, a plastic button, and an iron nail. He tried bringing a magnet close to each of these items.
In your opinion, which of these objects are attracted to the magnet? Explain your reasoning and give one other example of an object around you that is also attracted to magnets.

Figure 5. Flexibility Indicator

Item 6
Students are able to fluently and diversely mention various examples of the effects of force on objects in everyday life.

Instructions:

- Read the question carefully.
- Think about activities you often do at home or school.
- Choose three activities that involve pushing or pulling forces.
- Explain how the push or pull force affects the object involved in each activity.
- Write your answer clearly and completely.

Question;
Think about activities you usually do at home or at school. Write down three examples of activities that involve pushing or pulling forces. Explain how the force affects the object in each activity.

Figure 6. Fluency Indicator

Item 7
Students are able to explain in detail how force affects the motion and shape of objects under various conditions.

Instructions:

- Read the question carefully.
- Understand the phenomenon that occurs when you press a rubber ball.
- Explain why the rubber ball returns to its original shape after being released.
- Provide one other example of an object around you whose shape can also change due to force.
- Write your answer clearly and completely.

Question:

When you press a rubber ball, it changes shape. However, after being released, the ball returns to its original shape. Explain why this happens. Give one more example of an object around you whose shape can also change due to force!

Figure 7. Elaboration Indicator

Item 8
Students are able to analyze the relationship between types of landforms and appropriate types of jobs or professions.

Instructions:

- Read the question carefully.
- Choose one type of landform (rice field, mountain, or beach).
- Write the type of job that you think can be done in that place.
- Also explain why the job is suitable for that landform.
- You may write about a common job or one that is not often thought of.
- Use your own words and explain as clearly as possible.

Question:

Observe the following three types of landforms: rice field, mountain, and beach.
Choose one of these landforms.
Describe a job that you think people can do in that place.
Your answer may describe a common job or one that is rarely mentioned.
Also explain why the job is suitable to be done in that place!

Figure 8. Originality Indicator

Based on the results of the instrument test for Item No. 5, which assesses the flexibility indicator of creative thinking skills, 3 students scored 4, 18 students scored 3, 10 students scored 2, and no students scored 1, with a total of 31 students participating in the instrument trial. For Item No. 6, which measures the fluency indicator, 1 student scored 4, 11 students scored 3, 15 students scored 2, and 4 students scored 1. In Item No. 7, which targets the elaboration indicator, 1 student scored 4, 6 students scored 3, 15 students scored 2, and 9 students scored 1. Lastly, for Item No. 8, which focuses on the originality indicator, no students scored 4, 4 students scored 3, 13 students scored 2, and 14 students scored 1. The table of creative thinking indicators above presents the distribution of students' scores based on the tested indicators: flexibility, fluency, elaboration, and originality. The analysis of score distribution was limited to these four creative thinking indicators.

The result of this study is a HOTS (Higher Order Thinking Skills) instrument designed to measure students' critical and creative thinking abilities. The analysis of the findings in relation to the research questions in this developmental study is described as follows.

a. Validity of the HOTS Instrument

Critical and creative thinking are fundamental cognitive abilities that need to be nurtured from an early age, as they shape students' problem solving, decision making, and innovation capacity, as emphasized by Muid et al. (2024). Therefore, developing a valid and reliable assessment instrument to measure these skills in elementary school students is essential for supporting effective learning processes and meaningful evaluation.

The validity of the HOTS instrument developed to assess students' critical and creative thinking skills was evaluated through two stages: expert validation (content validity) and empirical testing (empirical validity through statistical analysis). Expert validation was conducted by three experts namely a subject matter expert, an evaluation expert, and a language expert. The results of the content validation revealed an average validity score of 0.82, which falls into the category of highly valid. This finding aligns with Ihsan (2015), who explains that an instrument can be considered valid when its content validity index reaches a high category, typically with a value above 0.80.

Empirical testing was carried out by calculating reliability using Cronbach's Alpha coefficient. Out of the 10 essay items developed, 8 items were found to be valid, while 2 items were declared invalid based on statistical analysis. According to Fitriana (2022), an assessment instrument can be considered reliable and valid if it meets the required validity criteria and obtains a Cronbach's Alpha coefficient greater than 0.70. Therefore, the results of this study indicate that the instrument has good reliability.

The importance of expert involvement during the content validation process is further emphasized by Nabil et al. (2022), who argue that expert judgments through procedures such as Aiken's V analysis significantly enhance the quality, relevance, and clarity of an assessment instrument. In addition, Musfirah et al. (2025) emphasize that empirical testing of assessment instruments is essential to identify items that do not function optimally, thereby improving the overall precision and effectiveness of the instrument.

The validity and reliability findings of this study are also in line with recent empirical research. Yulianti and Herpratiwi (2024) developed a SETS-based IPAS module (Science, Environment, Technology, and Society) that significantly improved elementary students' critical thinking skills, while a follow-up study involving a problem-based e-module also demonstrated enhanced critical engagement through integrated IPAS content. Similarly, Azharini et al. (2023) conducted a needs analysis and developed e-assessment tools oriented to Higher Order Thinking Skills (HOTS) in thematic learning, which were designed to measure both soft skills and hard skills, including critical and creative thinking abilities.

Supporting these findings, Srirahayu and Sulistyo (2018) emphasize that combining content validity procedures with empirical validation produces assessment instruments that are both accurate and well targeted in measuring students' competencies. Likewise, research by Trimawati et al. (2020) shows that assessment instruments developed in project-based

learning contexts are considered feasible when they meet key criteria such as validity, reliability, and practicality; in their study, instruments designed to assess critical and creative thinking in integrated science learning demonstrated high validity and very high practicality.

Based on these results, it can be concluded that the final product the HOTS instrument developed in this study meets the essential criteria of a proper assessment tool, namely validity and reliability. Thus, it is deemed suitable for measuring elementary school students' critical and creative thinking skills.

b. Practicality of the HOTS Instrument

The practicality test in this study was conducted through assessments by three educators using a questionnaire, which was analyzed quantitatively in a descriptive manner. The results showed that the average practicality score obtained was 93.6%, which falls into the "very practical" category. This indicates that the developed assessment instrument possesses a high level of usability for educators in the learning context.

These findings are consistent with research by Antari (2022), who reported very high practicality ratings from teachers and students in the use of HOTS-based assessment instruments. In line with common interpretation guidelines, practicality scores above 85% are generally categorized as "very practical" for instructional instruments. This suggests that the instrument is easy to use, easy to understand, and appropriate for use in elementary school settings. In addition, Anselmus, Risalah, and Sandie (2021) emphasize that involving teachers in the practicality testing process is essential to reflect the actual implementation of the instrument in the field. In this context, the educators' assessments provide direct insight into the extent to which the instrument can be used without significant obstacles.

The practicality of the instrument can be observed through its ease of use, time efficiency, and clarity of instructions, all of which are emphasized as key indicators of a practical assessment tool by Hernawan (2018). The instrument was designed with clear guidelines, a simple format, and an easily understood scoring mechanism. A practical assessment instrument not only makes it easier for teachers to use in classroom evaluation but also supports more effective learning, particularly in measuring higher-order thinking skills such as critical and creative thinking (Antari, 2022). This aligns with the context of the HOTS instrument developed in this study.

Similar results were also found in a study by Setiani (2022), which showed that HOTS-based assessment instruments with a high level of practicality can be implemented effectively in elementary schools and are feasible to use in classroom learning. This is highly relevant to conditions in elementary schools, where teachers often face limitations in time and resources for applying complex assessment tools. The practicality assessment in this study is also aligned with the principles of practicality outlined by Fernando and Sarkity (2022), as well as the view of Fitriana (2022) that assessment instruments must meet validity and reliability requirements before they are feasible for use. In this study, a practical assessment instrument is interpreted as fulfilling several key aspects, namely: (1) alignment with assessment objectives; (2) simplicity in usage and interpretation; (3) reliability of results; and (4) ease of scoring guidelines.

By meeting all of these aspects, the HOTS instrument developed in this study can be concluded to be highly practical and suitable for use by educators in the teaching and evaluation of students' critical and creative thinking skills.

c. Readability of the HOTS Instrument

The readability test was conducted by involving 10 fourth-grade students as respondents. Data analysis using a quantitative descriptive approach showed that the readability level of the instrument achieved an average score of 84.58%, which falls into the "very good" category. This result indicates that, in terms of language, question instructions, and sentence structure, the instrument was considered easy to understand by students. The high level of readability suggests that the instrument has been appropriately adjusted to the cognitive development and linguistic abilities of fourth-grade students, thereby minimizing the risk of misinterpretation. This finding is consistent with Triyoso et al. (2017), who state that assessment instruments with good readability are closely aligned with students' language abilities and cognitive development levels.

The use of simple, systematic, and contextually appropriate sentences contributes significantly to students' ability to comprehend and respond to the items correctly (Trimawati et al., 2020). Furthermore, Nurmalasari (2023) highlights that readability is a crucial aspect that supports the feasibility and appropriateness of assessment instruments, especially when they are used with elementary school students. High readability enhances the accuracy of students' responses and supports the effective implementation of assessments, as shown by the findings of Sito, Galib, and Sukariasih (2024). This is further supported by Puspita and dewi (2021), who emphasize that assessment instruments should be aligned with children's developmental stages to prevent ambiguity

and to ensure that learning outcomes are measured accurately.

Based on the distribution of scores from the instrument trial, it was found that most students tended to demonstrate stronger critical thinking skills compared to creative thinking. This is indicated by the higher number of students achieving high scores (scores of 3 and 4) on critical thinking indicators such as interpretation, analysis, and evaluation, in contrast to creative thinking indicators such as flexibility, fluency, elaboration, and originality, which tended to show lower score distributions (scores of 1 and 2). These findings suggest that, in general, students exhibit stronger performance in critical thinking than in creative thinking.

The pilot test of the HOTS instrument also revealed variations in students' critical and creative thinking abilities. For critical thinking indicators such as interpretation, analysis, and evaluation, the majority of students scored in the medium category, with only a few reaching the high category. Among the creative thinking indicators, flexibility showed relatively better results compared to fluency, elaboration, and originality, which mostly fell into the lower-to-medium category. These findings indicate that the developed instrument was effective in distinguishing students' levels of higher order thinking and in identifying areas that require further development.

These results are consistent with the theory of Anderson and Krathwohl (2001), which states that a good assessment instrument should be able to reveal variations in higher order thinking skills, including both critical and creative thinking. Furthermore, the findings align with the classical theory of Torrance (1974), who emphasized that aspects of creativity such as flexibility and elaboration deserve particular attention due to their significant potential for development in elementary school students

Conclusion

This study successfully developed a HOTS assessment instrument to measure the critical and creative thinking abilities of fourth-grade elementary school students. The instrument was proven to be valid (0.82), reliable (α = 0.865), practical (93.6%), and readable (84.58%), fulfilling essential criteria for a proper assessment tool. Results showed that students' critical thinking skills were generally stronger than their creative thinking skills, particularly in flexibility compared to fluency, elaboration, and originality. Thus, the developed instrument is effective for identifying students' higher-order thinking skills and can be applied to support meaningful learning and evaluation in elementary education.

Acknowledgments

The researcher sincerely expresses gratitude to the expert lecturers from the University of Lampung and to the practicing teachers for their valuable assistance and contributions in the validation process of the developed instrument. Appreciation is also extended to SDN 01 Bakung Udik and SDN 01 Bakung Ilir for granting permission and support during the implementation of this research, as well as to the beloved fourth-grade students who participated in the study.

Author Contributions

Conceptualization, J.S., U.R., and D.Y.; methodology, D.Y., and M.W.; software, U.R., and D.Y.; validation, U.R., and M.W.; investigation, J.S., and U.R.; writing original draft preparation, writing, J.S.; review and editing, U.R., D.Y., and MW.

Funding

The funding for this article was sourced from the researcher's personal funds.

Conflicts of Interest

There are no known conflicts of interest associated with this publication.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman. Retrieved from https://openlibrary.org/books/OL18156636M/A_taxonomy_for_learning_teaching_and_assessing

Antari, N. M. D. (2022). *Pengembangan instrumen tes IPA high order thinking skills (HOTS) untuk siswa SMP kelas VIII pada materi getaran, gelombang dan bunyi* (Skripsi sarjana, Universitas Pendidikan Ganesha). Retrieved from http://repo.undiksha.ac.id/id/eprint/13356

Anselmus, A., Risalah, D., & Sandie, S. (2021). Pengembangan instrumen tes HOTS dalam mengukur kemampuan berpikir analitis siswa pada materi program linear di kelas XI SMA Sungai Kehidupan. *Juring (Journal for Research in Mathematics Learning)*, *4*(4), 371–378. https://doi.org/10.24014/juring.v4i4.14295

Azharini, R., Handoko, H., Rosidin, U., & Perdana, R. (2023). Analysis of student needs for e-assessment tools oriented to higher order thinking skills (HOTS) in thematic learning to measure soft skills and hard skills competence. In R. Perdana et al. (Eds.), *Proceedings of the 4th International Conference on Progressive Education 2022 (ICOPE 2022)* (pp. 79–87). Atlantis Press. https://doi.org/10.2991/978-2-38476-060-2_9

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria, VA: ASCD. Retrieved from https://books.google.co.id/books/about/How_to_Assess_Higher_order_Thinking_Skil.html?id=AFIxeGsV6SMC

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi report)*. Millbrae, CA: California Academic Press. Retrieved from https://catalog.hathitrust.org/Record/009922105

Fernando, A., & Sarkity, D. (2022). Pengembangan instrumen uji validitas dan praktikalitas media pembelajaran IPA. *Pedagogi Hayati: Jurnal Ilmiah Pendidikan Biologi, 6*(2), 67–77. https://doi.org/10.31629/ph.v6i2.5212

Fitriana, D. A. (2022). Validitas dan reliabilitas instrumen penilaian matematika dengan problem solving berbasis penskoran politomus. *IKLILA: Jurnal Studi Islam dan Sosial, 5*(1), 13–25. Retrieved from https://ejournal.iaikhozin.ac.id/ojs/index.php/iklila/article/download/118/91

Hernawan, A. (2018). *Pengembangan instrumen asesmen berpikir kritis pada pembelajaran tematik kelas IV sekolah dasar di Kecamatan Tumijajar* (Tesis magister, Universitas Lampung). Retrieved from https://digilib.unila.ac.id/30189/

Ihsan, H. (2015). Validitas isi alat ukur penelitian: Konsep dan panduan penilaiannya. *Pedagogia: Jurnal Ilmu Pendidikan, 13*(2), 266–273. https://doi.org/10.17509/pedagogia.v13i2.3557

Karim, K., & Normaya, N. (2015). Kemampuan berpikir kritis siswa dalam pembelajaran matematika dengan menggunakan model Jucama di sekolah menengah pertama. *EDU-MAT: Jurnal Pendidikan Matematika, 3*(1), 92–104. Retrieved from https://ppjp.ulm.ac.id/journal/index.php/edumat/article/view/634

Kusuma, M. D., Rosidin, U., Abdurrahman, A., & Suyatna, A. (2017). The development of higher order thinking skill (HOTS) instrument assessment in physics study. *IOSR Journal of Research & Method in Education, 7*(1), 26–32. Retrieved from https://www.iosrjournals.org/iosr-jrme/papers/Vol-7%20Issue-1/Version-5/D0701052632.pdf

Muid, A., Albab, M. U., & Ibad, M. A. (2024). Pengembangan instrumen evaluasi. *Jurnal Ilmu Pengetahuan dan Pendidikan Islam, 14*(14). Retrieved from https://jurnal.maziyatulilmi.com/index.php/jippi/article/view/109

Munandar, U. (2017). *Pengembangan kreativitas anak berbakat* (Cet. ke-3). Jakarta, Indonesia: Rineka Cipta. Retrieved from https://library.ulb.ac.id/index.php?id=21783&p=show_detail&utm

Musfirah, N., Nurbaya, N., Nursalam, N., & Rasyid, A. (2025). Pengembangan instrumen hasil penilaian belajar tes dan non tes. *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial, 2*(11), 120–125. https://doi.org/10.5281/zenodo.15534085

Nabil, N. R. A., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *Paedagogia, 25*(2), 184–191. https://doi.org/10.20961/paedagogia.v25i2.64566

Nurmalasari, W. (2023). Penilaian autentik pada pembelajaran membaca pemahaman siswa kelas IV sekolah dasar. *Jurnal Pendidikan Dewantara: Media Komunikasi, Kreasi dan Inovasi Ilmiah Pendidikan, 9*(1), 30–43. https://doi.org/10.55933/jpd.v9i1.488

Puspita, V., & Dewi, I. P. (2021). Efektifitas E-LKPD berbasis pendekatan investigasi terhadap kemampuan berfikir kritis siswa sekolah dasar. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 5*(1), 86–96. https://doi.org/10.31004/cendekia.v5i1.456

Setiani, A. E. (2022). Pengembangan instrumen penilaian "HOTS" dari buku tematik dengan menggunakan Quizizz di sekolah dasar. *el-Ibtidaiy: Journal of Primary Education, 5*(1), 90–103. https://doi.org/10.24014/ejpe.v5i1.14324

Sito, L., Galib, L. M., & Sukariasih, L. (2024). Pengembangan instrumen asesmen kompetensi pada materi pesawat sederhana untuk peserta didik kelas VIII sekolah menengah pertama. *Jurnal Penelitian Pendidikan Fisika, 9*(1), 44–56. https://doi.org/10.36709/jipfi.v9i1.106

Srirahayu, R. R. Y., & Sulistyo, I. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian dan Evaluasi Pendidikan, 22*(2), 178–190. https://doi.org/10.21831/pep.v22i2.20270

Torrance, E. P. (1974). *Torrance tests of creative thinking: Norms-technical manual*. Lexington, MA: Personnel Press. Retrieved from https://openlibrary.org/books/OL14637755M/Torrance_tests_of_creative_thinking

Triyoso, A., Ibrahim, M., & Raharjo, R. (2017). Pengembangan instrumen asesmen komprehensif berbasis kompetensi pada mata pelajaran IPA/Biologi. *JPPS (Jurnal Penelitian Pendidikan Sains), 2*(2), 258–263. https://doi.org/10.26740/jpps.v2n2.p258-263

Trimawati, K., Kirana, T., & Raharjo, R. (2020). Pengembangan instrumen penilaian IPA terpadu

dalam pembelajaran model Project Based Learning (PjBL) untuk meningkatkan kemampuan berpikir kritis dan kreatif siswa SMP. *Quantum: Jurnal Inovasi Pendidikan Sains, 11*(1), 36–52. https://doi.org/10.20527/quantum.v11i1.7606

Yulianti, D., & Herpratiwi, H. (2024). Development of a science, environment, technology, and society-based learning module to foster critical thinking in elementary students. *Journal of Education and Learning (EduLearn), 18*(4), 1372–1384. https://doi.org/10.11591/edulearn.v18i4.21713