Research Article

Development of Critical Thinking Test Instruments for Elementary School Students in Ambon City

Andrea Mawar Hutasoit, Cristina Ginting, Haganta Tarigan, and Marleny Leasa*

Department of Primary Teacher Education, Faculty of Teacher Training and Education, Pattimura University Ambon - Indonesia

ORCID
Marleny Leasa: https://orcid.org/0000-0003-0718-3447

Abstract.
The ability to think critically is a skill that must be possessed by elementary school students in the era of globalization. This ability enables learners to analyze information, solve problems, and make informed decisions. This study aimed to develop a valid and reliable critical thinking test instrument to measure the critical thinking ability of elementary school students in Ambon City. This study is a type of research and development with the stages: needs analysis, instrument development, instrument validation, and instrument testing. The subject of the study is a 5^{th} grader of public elementary school 1 Latihan SPG Ambon. This study produced a critical thinking test instrument consisting of 13 valid and reliable questions. The readability test results show that the validation of each item r count > r table and reliability of each item greater than > 0.60 are declared reliable. After the readability test, it was tested on a large scale. It was obtained that the critical thinking skills of elementary school students in 5^{th} grade in Ambon City are generally classified as medium, with the acquisition of an average value of 32%. The empirical test results show the reliability of the instrument of 0.941.

Keywords: critical thinking, science, test instruments

Corresponding Author: Marleny Leasa; email: marlenyleasa3@gmail.com

Published: 11 November 2024

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the 8th Isedu Conference Committee.

1. Introduction

The challenges of the 21st century require innovative solutions based on scientific thinking and important discoveries. The learning paradigm in the 21st century emphasizes that learners are the center of learning so that they can use various sources of information effectively, analyze information critically, present information creatively, and convey information independently, not just receive it from the teacher. The availability of various sources of information needs to be filtered so that the information taken can develop the potential of learners [1].

One aspect of the internationally competitive education scope is the level of critical thinking of students in the field of science for elementary school (SD). Based on research

conducted by the Center for Educational Assessment (Puspendi) at the Ministry of Education and Culture in 2021, the critical thinking skills of elementary school students are still relatively low. The results showed that only 25% of elementary school students have good critical thinking skills. The remaining 75% of elementary school students have moderate or low critical thinking skills.

The Indonesian government's policy to strengthen human resources in education is pursued by dynamizing the curriculum. According to Law No.20 of 2003, Chapter 1 Article 1 states that "curriculum is a set of plans and arrangements regarding the objectives, content, and learning materials as well as the methods used as guidelines for organizing learning activities to achieve certain educational goals." The Merdeka curriculum is an educational approach that aims to encourage students' independence and critical thinking skills [2].

In essence, the Merdeka Curriculum is an educational approach that aims to develop students' independence and critical thinking skills in a rapidly changing and complex world. The Merdeka Curriculum can positively impact the development of students' independence and critical thinking skills. This approach will help students face the rapidly changing world of work and global challenges and prepare them to become individuals who are independent, critical, and able to adapt to a dynamic environment [3].

This concept emerged in response to the changing needs of the world of education, where critical thinking skills and independence are increasingly crucial in helping students face future challenges. In the Merdeka Curriculum, education no longer only focuses on transferring knowledge from teachers to students but emphasizes student-centered learning. Students are free to explore their interests and talents and are encouraged to take responsibility for their learning. The teacher acts as a facilitator and guide, providing support and guidance in the learning process.

Teachers play an important role in fostering students' critical thinking skills by creating diverse opportunities and opportunities. These opportunities must be tailored to students' individual characteristics so that each student can develop optimally. Encouraging students to be more open and minimizing their anxiety is also important. According to [4], the main challenge in learning is how students can learn well by exploring their learning styles. Students who are independent and have high academic abilities have different ways of learning from those with limitations. Teachers must consider this when creating the right learning opportunities for each student. Thus, each student's critical thinking skills can develop and learn optimally.

Thinking skills that can be directed through science learning in elementary school are higher-level thinking skills. In general, skills consist of four levels: Recall thinking (memorization), basic Thinking, critical thinking, and creative thinking [5]. Critical thinking skills can empower students to develop themselves into confident and independent people who can solve problems in learning in the classroom and society [6].

Critical thinking skills are not acquired automatically but need to be practiced [7]. Thinking is a natural process; it can lead to misperceptions and a lack of information without practice [8]. Learners can improve their thinking skills by thinking critically, such as evaluating the information used. Therefore, although learners have a natural ability to think critically, teacher guidance is essential to help them improve these skills. Learning is one of many ways to develop critical thinking skills. Proper evaluation can also improve it [9]. Tests with indicators such as asking and answering questions with explanations, deduction, induction, decision judgment, and taking action can improve one's critical thinking skills.

Although education has emphasized the development of critical thinking skills, assessment instruments to measure them still need to catch up [10]. Appropriate measurement tools must support efforts to assess learners' success in developing this ability. This measuring instrument must represent students' critical thinking skills, such as critical thinking test questions specifically designed to measure these skills.

Classical methods still dominate learning, and teachers also need help making test instruments to measure students' critical thinking skills. The assessment only measures the level of memorization knowledge, not the ability to think critically [11]. Schools rarely use standardized tests because although standardized tests are better than teacher-made tests, the number in the world of education is still scarce. It causes less accurate teacher assessment of student's ability to understand the material.

Teachers use essay assessment instruments and multiple-choice tests available in student books, only measuring students' memorization and understanding. Multiple choice tests are often used because they can measure more material/competencies to be measured, are more efficient in assessing, and students are more accessible. The limited number of test instruments to measure elementary school students' critical thinking skills is an obstacle in the learning evaluation process.

Researchers use instruments to collect data by making measurements [12]. Critical thinking test instruments must be developed in elementary schools because critical thinking skills are essential to master early on. Developing critical thinking test instruments in elementary schools can assist teachers in evaluating the effectiveness of learning and designing more effective learning strategies. By having a valid and reliable

test instrument, teachers can measure students' progress in developing critical thinking skills and identify areas where students need additional help. This allows teachers to customize their learning approach according to students' needs. In this study, the development is to produce a critical thinking test instrument for blood circulation system material.

There are many benefits from developing critical thinking test instruments, one of which is to determine the ability of critical thinking skills of elementary school students. It will help students hone their potential [13] so that they can prepare themselves to face challenges and developmental tasks [14]. The teaching and learning process trains students to think so that the ability to think will produce intelligent students who can solve every problem they face. In that way, students can see the extent of their abilities.

Developing a critical thinking test instrument involves several steps that need to be followed. The following are the steps:

1. Identify the components of critical thinking
2. Developing test questions
3. Testing the test instrument
4. Analyzing data
5. Testing Instrument Revision

Developing critical thinking test instruments for elementary school students is an essential step in improving students' critical thinking skills. So, this research aims to produce an instrument for testing elementary students' critical thinking skills.

2. Method

This research uses a type of development research using the ADDIE development model [15] the ADDIE model is one of the basic models in development research.

2.1. Analysis

In analysis, researchers collect data related to a problem in the field. The needs analysis in this study (1) conducted a literature review on the concepts, theories, and indicators of critical thinking skills of elementary school students, (2) reviewed the elementary school curriculum to find out learning objectives related to critical thinking skills, (3) conducted observations and interviews with elementary school teachers and students to find out and problems in learning critical thinking. Objective analysis formulates the

development of measurable critical thinking test instruments for elementary school students that meet their needs. Target analysis determines the target group for using the test instrument: elementary school students in certain classes or throughout Ambon.

2.2. Design

In this study (1) the preparation of instrument blueprint, namely to develop a blueprint of test instruments that contain aspects, indicators, and items that are in accordance with the results of the objective needs analysis. (2) Preparing and compiling test items that meet the rules of writing good questions.

2.3. Development

At the development stage, something is needed to develop research that will be carried out so that the products needed in the research are produced. The development stage is the preparation of instrument prototypes, instrument validation, and instrument trials.

2.4. Implementation

The implementation stage is the testing stage after the product has been produced so that the validity, reliability, and usefulness can be measured and tested. The critical thinking skills test of elementary school students was carried out using the instrument that had been developed.

2.5. Evaluation

The evaluation stage can be done as well as the previous stages by providing formative or summative evaluations so that students gain knowledge and understanding during the learning process.

3. Results and Discussion

3.1. Results of Theoretical Instrument validation

The validation process of the cognitive test instrument begins with giving the blueprint along with the content validation sheet to two experts. Based on Gregory's analysis of the results of validation by experts, it can be seen that 15 items are suitable for testing. The results of theoretical validation by experts are suitable for testing without revision.

3.2. Empirical Instrument Validation Results

Empirical validation was done by testing 13 valid items and two invalid items. By knowing the validity of a question item, its reliability can be known. The reliability is 0.914 with a total of 13 question items. In addition to item testing and instrument reliability, quantitative item analysis is also carried out at the item difficulty level, differentiating power. The following are the results of analyzing the items' difficulty levels.

TABLE 1: Problem Difficulty Test Results.

Item Problem		1	2	3	4	5	6	7	8	9	10	11	12	13
N	Valid	25	25	25	25	25	25	25	25	25	25	25	25	25
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean		.8400	.8800	.6400	.5600	.7200	.7200	.5200	.2400	.4400	.4400	.3200	.2800	.2800

Based on the table above, it can be explained that of the 13 questions, there are 3 difficult category questions, 6 medium category questions, and 4 easy category questions.

Description:

* Blue color Hard Problem

* Red color Medium Problem

* Black color Easy Problem

The results of the differentiating power analysis can be seen in the following table.

The test instrument is an essay question with a total of 15 items. Expert validation questionnaires regarding material, construction, language, and critical thinking skills have the first category, irrelevant (score 1) and less relevant (score 2), categorized into weak relevance, and the second category for moderately relevant (3) and highly relevant (4). The results of data analysis were used to process data from expert validation

Table 2: Differentiating Power Results.

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Category
Problem 1	6.04	48.207	.912	.894	Excellent
Problem 2	6.00	50.750	.808	.899	Excellent
Problem 3	6.24	52.023	.844	.898	Excellent
Problem 4	6.32	58.060	.431	.915	Good
Problem 5	6.16	50.057	.761	.902	Excellent
Problem 6	6.16	54.140	.690	.905	Excellent
Problem 7	6.36	50.907	.804	.899	Excellent
Problem 8	6.64	60.740	.431	.914	Good
Problem 9	6.44	58.923	.466	.913	Good
Problem 10	6.44	58.173	.538	.911	Good
Problem 11	6.56	60.423	.381	.915	Simply
Problem 12	6.60	57.000	.772	.905	Excellent
Problem 13	6.60	60.500	.443	.914	Good

regarding material, construction, and language using the Gregory test method [16]. The results of validation using the Gregory test are listed in Table 3.

Validity is the determination of the interpretation obtained from the assessment results. The validation of a test instrument can be interpreted as the ability of the test to measure what should be measured [17]. Content validity is a measurement of the extent to which an assessment instrument covers content that is relevant to the construct being measured, and ensures that its items present essential aspects of the construct. Content validity involves checking the suitability of test items made with indicators, materials, or learning objectives that have been applied. Content validity in an instrument describes the extent to which the instrument reflects the content of the subject matter being evaluated [18]. A test is considered to have content validity if it can measure specific objectives that align with the material or content of the lesson being taught [19].

Construct validity is the extent to which a measurement instrument measures the intended construct and can distinguish that construct from others [20]. Construct validity involves accumulating empirical evidence from various sources, including factor analysis, convergent and divergent relationships, and correlational research, to ensure that the measurement instrument effectively measures the construct under study. The

Table 3: Validation Questionnaire of the Test Instrument.

Aspects	Indicator	Validator score		Average
		1	2	
Material	Questions are fit for purpose (requires a written test for description questions and multiple choice questions for multiple choice questions)	4	4	4
	Question boundaries and expected answers are appropriate	4	4	4
	The stated material is by the competencies	4	4	4
	The content of the specified questions is appropriate for the educational level of the learners	4	4	4
	The specified question material is by the time allocation	4	4	4
Construction	The assessment instruments and guidelines are based on the chosen form and assessment technique.	4	3	3,5
	Uses question words or command words that require multiple choice and description answers	4	4	4
	There are clear instructions for doing the questions	3	3	3
	There is a rubric or guideline for giving an assessment score	4	4	4
	The scoring rubric comes with clearly outlined descriptors	4	3	3,5
Language	Formulation of commutative problem sentences	4	4	4
	Question items use standard Indonesian language	4	3	3,5
	Do not use words/expressions that give rise to multiple interpretations	3	3	3
	Not using local or taboo language	4	3	3,5
	Total Score	54	50	52
	Average Score	3,85	3,57	3,71

Judges I		Judges II	
Not Relevant (score 1-2)	Relevant (Scores 3-4)	Not Relevant (score 1-2)	Relevant (Scores 3-4)

construct validation process begins by identifying and delimiting the variables to be measured and then describing them in a logical structure based on related theory. From the theory, practical implications can be drawn regarding the measurement results under certain conditions, and these implications will be tested. If the results are as expected, the instrument will have appropriate construct validity.

The validation results in Table 1 show that the total score from Validator 1 is 54, and the score from Validator 2 is 50, with a total score of 52. Therefore, the average obtained from validator 1 is 3.85, and from validator 2 is 3.57. The overall average of the two validators is 3.71, categorized as quite relevant. In content validity, there are various ways that can be used, the purpose of which is to see the agreement of 2 or more experts. Content validity describes the extent to which the questions, tasks, or items in

a test or instrument are able to represent the overall and proportional behavior of the sample being evaluated.

Based on the results of the above assessment in Table 3, the validity statement is very valid. This category is obtained by analyzing the validity of experts who were previously entered into the 2x2 cross-tabulation column, as in Table 4.

TABLE 4: Test Instrument Validation Questionnaire.

Expert Validation Assessment Tabulation	Judge I		
		Not Relevant (Score 1-2)	Relevant (Scores 3-4)
Judges II	Not Relevant (Score 1-2)	A 0	B 0
	Relevant (Scores 3-4)	C 0	D 30

By analyzing validity by 2 experts using the Gregory Formula

$$V = \frac{D}{(A+B+C+D)}$$

$$V = \frac{30}{(0+0+0+14)}$$

$$V = \frac{30}{30}$$

$$V = 1$$

Where:

V = construct validation

A = Both raters disagree (Weakly)

B = Both raters disagree (Weak-strong)

C = Rater 1 disagrees, rater 2 agrees (Strong-weak)

D = Both raters agree (Strong-strong)

The criteria in content validity include 1) 0.8-1 = Very high validity, 2) 0.6 - 0.79 = High validity, 3) 0.40 - 0.59 = Medium validity, 4) 0.20 - 0.39 = Low validity and 0.00 - 0.19 = Very low validity. Data from Table 3 show that there are no categories of questions that are said to be Weak-Weak, Weak-Strong, or Strong-Weak. Among the 15 validation aspects above, both experts gave scores in the Strong-Strong category, so with the Gregory formula (2007), the validity result with 100/100 is 1, where 1 is a very high validity criterion. Therefore, the validity of the results of the question instrument on blood circulation system is very high.

4. Conclusions and Recommendations

The expert validity questionnaire's material, language, and construct aspects show that overall, the 15 items in the test instrument are categorized as quite relevant with an average score of 3.71. Of the 14 aspects of validation mentioned, both experts gave scores in the Strong-Strong category, so the validity results reached 30/30, which is equivalent to 1. The number 1 indicates a very high level of validity. Based on this, the question instrument on blood circulation system has very high validity, with a score between 0.8 to 1, which is categorized as very high validity.

References

[1] Virkus SW. 21st Century Learning Paradigm. A paper presented at the National Seminar on Science Education, University of Education Indonesia, Bandung. 2011.

[2] Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System. https://jdih.kemdikbud.go.id/sjdih/siperpu/dokumen/salinan/UU_tahun2003_nomor020.pdf

[3] Siswati NA. Development of students' independence and critical thinking skills through Merdeka Curriculum. J Early Childhood Islamic Educ. 2022;9(2):235–52.

[4] Leasa M, Corebima AD, Batlolona JR. The effect of learning styles on the critical thinking skills in natural science learning of elementary school students. Elementary Education Online. 2020;19(4):2086–97.

[5] Siswono TD. Developing higher order thinking skills through science learning in elementary school. J Sci Educ. 2009;2(2):127–38.

[6] Ennis RH, Sternberg RJ. Critical thinking in education. In: Eller JF, editor. Critical thinking: What is it and what can we do to teach it? Mahwah (NJ): Lawrence Erlbaum Associates; 2002. pp. 1–25.

[7] Duron R, Limbach B, Waugh W. Critical thinking framework for any discipline. Int J Teach Learn High Educ. 2006;17(2):160–6.

[8] Black S. Teaching students to think critically. The Education Digest, 70(6), 42-47. The effectiveness of problem-based learning with openended on dynamic electricity material on students' critical thinking ability. Thesis. Not Published. 2005.

[9] Benjamin R, Klein S, Steedle J, Zahneer D, Elliot S, Patterson J. The case for critical-thinking skills and performance assessment. Council for Aid to Education; 2016.

[10] Fitrifitanofa W, Waskito S, Budiharti R. Development of formative test instrument. J Phys Educ. 2013;1(1):90–110.

[11] Mustami. Research instruments: Concept and development. Yogyakarta: Student Library. 2015.

[12] Hasyda A, Arifin Z. Concept map based learning to improve critical thinking ability of Grade IV Students of SDN 1 Tanjung Batu, Sungai Raya District, Kubu Raya Regency [JIGSD]. Scientific Journal of Elementary Teachers. 2020;3(1):1–10.

[13] Hasyda S, Arifin. Application of problem-based learning model to improve critical thinking ability of Grade IV Students in Elementary School. Jurnal Pendas Mahakam. 2020;5(1):62–9.

[14] Cennamo K, Abell SK, Chung JH. The ADDIE instructional design model: A case study for training distance learners. Educ Technol Res Dev. 1996;44(3):31–53.

[15] Ratnawati H. Development of critical thinking ability test instrument for high school students on fluid dynamics material. J Sci Educ. 2016;5(2):127–34.

[16] Arifin Z. Education evaluation]. PT Bina Pustaka Sarana Semesta. 2017.

[17] DeVellis RF. A modern method for survey design. Sage Publications; 2016.

[18] Sugiyono D. Development research methods. Alfabeta; 2011.

[19] Trochim WM. Research methods: The basics. Sage Publications; 2008.