P-ISSN xxxx-xxxx || E-ISSN xxxx-xxxx
Indonesian Physics Education Research
Volume 5, No. 1, June 2024, pp. 52-66

Development of an Assessment Instrument for High School Students' Critical Thinking Skills in Sound Wave Material

Pepi Rahmayani and Wawan Bunawan
Faculty of Mathematics and Natural Science, State University of Medan
Pepirahmayani030597@gmail.com

Abstract

This research aims to innovate an assessment instrument for sound wave material and to develop a critical thinking skills assessment instrument that meets valid and reliable criteria. The research method used is the ADDIE model. The product development procedure follows the stages from analysis to implementation. The analysis phase, which is the preliminary stage, includes the analysis of existing types of instruments used in schools. The design procedure involves drafting the test blueprints and question formats. The development phase includes the creation of the instrument, validation by experts, limited trials, large-scale trials, and product implementation. The results of this study indicate that the type of assessment instrument developed is an essay analysis test and a problem-solving test oriented towards students' critical thinking skills, which was applied at SMA Negeri 2 Percut Sei Tuan. The instrument was found to be valid and reliable. It can measure the average student learning outcomes and assess the proportion of student mastery.

Keywords: *Instrument, Assessment, Critical Thinking Skills, ADDIE Method*

According to the National Education System Law No. 20 of 2003, education is a conscious and planned effort to create a learning environment and learning process so that students can actively develop their potential to possess spiritual strength, self-control, personality, intelligence, noble character, and the skills necessary for themselves, society, the nation, and the state (Hadijah, A., 2016). Education plays a crucial role in shaping a high-quality generation that will contribute to the nation's future. The development of science and the advancement of assessment instruments for learning outcomes significantly influence and play an essential role in the field of education.

Educational assessment, according to the Ministry of National Education Regulation No. 20 of 2007 on Education Assessment Standards, is the process of collecting and processing information to determine students' learning achievements. Van de Walle, as cited by Agus and Jailani (2014), stated that assessment principles and standards emphasize two key ideas: assessment should enhance students' learning and serve as a valuable tool for making instructional decisions. Assessment of learning outcomes by educators is conducted continuously to monitor students' learning progress and to improve the effectiveness of the learning process. Assessment of learning outcomes by educational institutions is carried out to evaluate students' achievement of competencies in all subjects. Assessment by the government is conducted in the form of national examinations, aiming to assess students' competency achievements at a national level in specific subjects within the science and technology group.

Assessment is not merely about collecting data on students but also about processing it to gain insights into the learning process and outcomes. It involves more than just administering tests; teachers must also follow up on the results for the benefit of the learning process.

To conduct assessments, teachers require assessment instruments in the form of questions that test cognitive, affective, and psychomotor abilities. Assessment is a crucial activity in physics education. It provides constructive feedback to both teachers and students, and the results can motivate students to achieve better (Agus Budiman & Jailani, 2014). Additionally, assessments can influence learning behavior, as students tend to direct their study activities toward the types of

assessments conducted by teachers. The quality of the assessment instruments directly impacts the accuracy of determining students' learning achievements. Therefore, assessment instruments play a strategic role in teachers' and schools' decision-making regarding students' learning outcomes, including higher-order thinking skills.

Physics is a branch of natural science (IPA) that plays a significant role in the development of assessment instruments for critical thinking skills. Integrated science education requires authentic assessments. Assessment is the process of collecting various data and information to provide a picture of students' learning progress. Teachers need to be aware of students' learning progress to ensure that they are undergoing the learning process correctly. The data and information collected through assessments are not intended to judge students' status. Proper science education should focus on helping students find their learning strategies rather than emphasizing how much information they acquire at the end of the learning period (Tanjung, 2017). Continuous assessment by educators aims to monitor the learning process and students' progress while improving the effectiveness of learning activities. Assessment by educational institutions evaluates students' competency achievements across all subjects. The actual goal of physics education requires the use of assessment instruments that not only cover memorization and understanding but also train critical thinking skills.

Natural science (IPA), including physics, is often less favored by students. Teachers play a vital role in determining the quality and quantity of the learning process. Therefore, teachers must carefully consider and select various teaching strategies and apply them according to the instructional goals, one of which involves using appropriate instruments or questions for students (Kopi, Kendek, & Ali, 2016).

Assessment instruments are an integral part of the assessment process in learning. Docktor and Heller, as cited by Amalia & Endang, stated that assessment plays a role in evaluating the learning process, progress, and outcomes. Assessment instruments include tests and grading systems, designed to assess students' understanding after studying a particular competency (Prasasti et al., 2012). Well-designed assessment instruments that align with the levels of thinking can enhance students' thinking abilities, particularly critical thinking. Developing critical thinking skills is essential because they are not innate (Redhana & Liliasari, 2008). Critical thinking education in schools, especially at the high school level, has not been adequately addressed, resulting in relatively low critical thinking skills among graduates. The low critical and creative thinking skills of graduates, from elementary to higher education in Indonesia, is often a concern (Reta, 2012).

Interviews with physics teachers at SMA Negeri 2 Percut Sei Tuan revealed that the assessment instruments currently used mainly measure memorization and understanding. Additionally, interviews with some students showed that they are less interested in studying physics because their teachers rarely use engaging media, instead opting for monotonous lectures, note-taking on the blackboard, and solving problems without fully conducting laboratory experiments. This approach has negatively impacted students' learning outcomes, with many students not meeting the school's minimum competency standards (KKM) of 70. The interviews further confirmed that the assessment instruments should include components that train students' critical thinking skills. Sound waves are a physics topic that requires both memorization and understanding. This topic is crucial as a foundational concept for students to grasp subsequent physics concepts. Therefore, research was conducted to develop an assessment instrument for critical thinking skills in the topic of sound waves. This topic was selected based on the indicators in the physics syllabus of the 2013 Curriculum.

Many studies have examined critical thinking skills, which require appropriate measurement tests, such as multiple-choice tests, skill tests (Ennis, 1993; Ennis, 1996), and essay

tests (Ennis, 1993). This study focuses on developing an assessment instrument for high school students' critical thinking skills. Previous research has also explored the development of critical thinking tests. Ennis (1993) developed a content-free critical thinking test. Additionally, other studies have developed critical thinking tests related to specific subjects, such as Kartimi and Liliasari (2012), who developed a multiple-choice critical thinking test in thermochemistry, and Amalia and Susilaningsih (2014), who developed an essay test on acids and bases.

Based on this research, an assessment instrument for high school students' critical thinking skills in sound waves will be developed. This development aims to evaluate the critical thinking abilities of students in learning the topic of sound waves in grade XI at SMA Negeri 2 Percut Sei Tuan.

Research Methodology

The research conducted is research and development using the ADDIE method (Analysis, Design, Development, Implementation, Evaluation). The population for this research consists of all students in grades XI and XII at SMA Negeri 2 Percut Sei Tuan for the 2018/2019 academic year. The sample for the pre-trial class, used for instrument validation testing, consists of 35 students from class XI MIA 3 for the 2018/2019 academic year. The sample for the limited-scale trial consists of 31 students from classes XI MIA 2 and XI MIA 2. The sample for the large-scale trial during the implementation phase consists of 35 students from class XI MIA 1. The research design follows the ADDIE method:

1. designing test instruments with sound wave indicators based on the school's syllabus, lesson plans, and curriculum;
2. designing test instruments with Ennis's 1985 critical thinking skill indicators;
3. communicating the design of critical thinking assessment instruments based on content indicators and Ennis's 1985 critical thinking skill indicators;
4. conducting discussions and evaluations of the critical thinking assessment instruments with a team of validators until the items are valid and suitable for use;
5. administering the validated assessment instruments to the sample as an implementation test for assessing high school students' critical thinking skills.

RESULT AND DISCUSSION

Research Result

1. Data Analysis of Direct Observation

Field observations were conducted by the researcher for approximately three months before the study. The results of these field observations, obtained through interviews and direct observations, include the following:

Table 1: Data on Reasons for Development Based on Field Observations

Reasons of development	Field research
types of assessment instruments in schools.	The instrument consists of multiple-choice tests with 15-40 questions, and short answer questions with 3-5 questions. The aspects measured are memorization and understanding of concepts.
school facilities and infrastructure	Based on observations, SMA Negeri 2 Percut Sei Tuan has spacious classrooms and media such as projectors (infocus) and laboratories.

Physics learning process	Based on documentation studies, the discussion activities conducted in the classroom are less conducive, and the approach used is conceptual.
Curriculum	The curriculum in use at SMA Negeri 2 Percut Sei Tuan is the 2013 Curriculum. However, the actual implementation does not reflect the 2013 Curriculum in the classroom teaching process.

2. Analysis of Physics Material Selection

The selection of physics material was conducted by the researcher in discussion with a physics subject teacher at SMA Negeri 2 Percut Sei Tuan. The results of the discussion are as follows:

Table 2: Results of Physics Material Selection Analysis

Komponen	Discussion Conclusion
Optic Material	22April 2019
Test Form	Tes Essai and Tes *Problem solving*
Total Test	≤ 20 question

3. Analysis of Material Qualification

The qualification of materials in this study involves selecting the focus of the material from the topic "Sound Waves." At this stage, the researcher conducted discussions with a lecturer, resulting in the following information:

Table 3: Results of Material Qualification Analysis

Komponen	Kesimpulan Diskusi
	Focus of the Material:
Optic Subject	Sound Waves, String Sound Sources, and Sound Propagation and the Doppler Effect
Essay test	15 test items cover the focus of the material.
Essat test	The test items include 10 essay analysis questions and 5 problem-solving questions, covering the focus of the material.

4. Instrument Preparation Analysis

The preparation of the test instruments for assessing Critical Thinking Skills in this study involved developing indicators for the Critical Thinking Skills instrument based on Ennis's 1985 indicators, using the Lesson Plan (RPP) and the implemented curriculum. This preparation aimed to establish a framework for determining the achievement distribution of each item in the critical thinking skills instrument.

5. Literature Review Analysis for Product Assessment

The literature review was conducted by gathering information on the characteristics of critical thinking skill questions used in schools. The literature collected included books, journals, and participation in seminars that support the development of critical thinking skill instruments.

6. Research Method Analysis

The selection of the research method was based on the analysis of data from field observations and literature reviews. Field observation data indicated that teachers often use test items provided in school textbooks, which predominantly cover the cognitive domains C1-C3, assessing only memorization and comprehension. The test questions used do not differ significantly from those used in previous years, with teachers often modifying only the numbers in the items given to students. Due to the limited research period of approximately three months, the researcher chose to use the ADDIE model for research and development.

According to Haya, Frilisa Dliyaul, et al. (2014:12), the ADDIE model, developed by Dick & Carey (1996) and modified by Molenda, stands for Analysis, Design, Development or Production, Implementation, and Evaluation. Based on the field data and the suitability of the Research and Development (R&D) ADDIE model, the researcher decided to use this method for the study.

7. Design Phase

Development of Critical Thinking Skills Assessment Instruments Based on Ennis's 1985 Indicators

The design of the Critical Thinking Skills assessment instrument involves developing test items through communication between the specified sound wave material and the established indicators, as well as the type of questions based on Ennis's 1985 indicators.

8. Design of the Assessment Used for Critical Thinking Skills Instruments

Assessment instruments are an integral part of the evaluation process in education. Docktor and Heller, as cited by Amalia & Endang, state that assessment plays a role in evaluating the process, progress, and outcomes of student learning. Assessment instruments include tests and grading systems designed to determine students' understanding after studying a particular competency (Prasasti et al., 2012).

Well-designed assessment instruments that align with levels of thinking can enhance students' cognitive abilities, particularly critical thinking. Critical thinking skills are essential to develop because they are not innate (Redhana & Liliasari, 2008). Currently, critical thinking education in schools, especially at the high school level, is not adequately addressed, resulting in relatively low critical thinking skills among graduates. The low critical and creative thinking

skills of graduates from elementary to higher education in Indonesia are often a concern (Reta, 2012).

9. Design of the Relevance of Critical Thinking Skills Instruments with Theory

The relevance of the Critical Thinking Skills instrument to theory is assessed by aligning the developed test with theories of Critical Thinking Skills as proposed by experts. The researcher collected information by comparing the test items with the characteristics of test items described in the book *"Keterampilan Berpikir (Pedoman dan Acuan Para Peneliti Keterampilan Berpikir)"* by Rusyna Adun and the critical thinking book by Surip.

10. Design of the Implementation of Critical Thinking Skills Instruments

The application of this instrument involves integrating all the results of the activities into a coherent framework for developing the test instrument. This application represents the conclusions drawn from each stage of comparison, including the alignment of item indicators with the material, and the final conclusions derived from comparing the material with the accuracy of the Critical Thinking Skills instrument based on Ennis's 1985 indicators and expert opinions.

11. Creating Critical Thinking Skills Instruments

At this stage, the test instruments have been fully developed. The results of the development of the Critical Thinking Skills instruments include 10 essay analysis questions and 5 problem-solving questions.

12. Conducting Content Validation of Critical Thinking Skills Instruments

Validation was carried out by 3 experts to validate the test instruments before they are applied to the research samples. In this stage, the validators provide suggestions and feedback to further improve the test instruments developed by the researcher.

13. Analysis of Content Validation and Reliability of Critical Thinking Skills Instruments

The developed test instruments are presented in two types: 10 essay analysis questions and 5 problem-solving questions. Before testing the developed instruments, the researcher first conducted validation testing with 3 experts. This validation analysis was performed using Microsoft Excel, employing the Product Moment Pearson correlation statistics.

From the data above, it can be concluded that the opinions of the three experts in the test instrument validation process were highly accurate.

The analysis produced the following data conclusions:

A child experiences things that they often hear but are confused about why they happen, such as: the echoing sound or reverberation in a closed space, the sound of thunder being louder at night compared to during the day, and hearing the sound of a car engine around a bend even though the car is not visible due to tall buildings blocking the view. Why does this not happen all the time? (A) Based on the above explanations, what can you infer about the phenomena experienced by the child? Are these phenomena related to the properties of waves in their everyday life? If so, why do you say that, and explain the properties of waves and their applications? (B) What is the intensity of human hearing at night?	• The form of the questions does not align with the answers and the indicators of students' critical thinking.	This happens because, at night, the temperature drops, which increases the density of the air and subsequently affects the speed of sound in the air. The lower the air temperature, the faster the speed of sound. Therefore, at night, we can hear things more clearly, including thunder, compared to during the day. The problem the child is experiencing is that they do not understand the relationship between air and the speed of sound, which makes it confusing for them to deal with this issue every day.
A sound source travels through a medium with a speed of 330 m/s. If the frequency of the sound is 200 Hz, what is the wavelength of the sound?	• The questions tested do not enhance students' thinking skills.	Known ; v = 330 m/s f = 200 Hz Question : λ = ? Answer ; $\lambda = \frac{v}{f}$ $\lambda = \frac{330\ ms^{-1}}{200\ Hz}$ λ = 1,65 m
A child hears the sound of thunder 1/16 of a minute after the lightning strike. If the speed of sound in air is 340 m/s, determine the distance from the lightning strike to the child.	• The form of the questions and the indicators used are not coherent.	Known ; v = 330 m/s F = 200 Hz Question : s....? Answer : s = v.t S = 3,75 x 340ms^{-1} S=1,275m
Validator	Skor	Criteria
Validator 1	3,58	Good
Validator 2	3,9	Good
Validator 3	3,94	Good

14. Implementation Phase

Data from Pre-Test Samples

Essay Analysis Test

a) Correlation Analysis (Item Validation)

Validiy	t Count	0,62	0,44	0,74	0,36	0,39	0,44	0,62	0,55	0,74	0,39
	t Table	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33
	V/T	V	V	V	V	V	V	V	V	V	V
Reliability	Variance	18,4	6,3	22,1	19,6	14,6	6,4	7,2	6,8	12,9	13,1
	J.Variance	128,0									
	Total Variance	307,5									
	Reliability	0,71									
Level of Difficulty	Average	3,57	3,71	3,85	7,42	8,28	9,08	6,88	9,02	4,14	8,28
	Level	0,35	0,37	0,38	0,74	0,828	0,90	0,68	0,90	0,41	0,82
	Criteria	Medium	Medium	Medium	Easy	Easy	Easy	Medium	Easy	Medium	Easy

Validiy	t Count	0,470521	0,38512	0,734247	0,347733	0,46515	0,774712061	0,780413	0,7795	0,53443	0,784859
	t Table	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334	0,334
	V/T	V	V	V	V	V	V	V	V	V	V
Reliability	Variance	58,25228	15,82857	0,843697	6,196639	1,470588	6,492436975	3,020168	4,398319	8,870766	6,596639
	J.Variance	176,78487									
	Total Variance	307,51									
	Reliability	0,74499	5	10	10	10	10	7	10	0	10
	Average	4,3714	6,2286	4,7429	4,4571	5,0000	9,0857	6,2571	9,3143	3,9118	9,1429
	Level	0,437143	0,622857	0,474286	0,445714	0,5	0,908571429	0,625714	0,931429	0,391176	0,914286

Level of Difficulty	Criteria	Medium	Medium	Medium	Medium	Medium	Easy	Easy	Easy	Easy	Easy

This data was then compared with the pre-trial data from the limited scale trials and the analysis data from the large-scale implementation phase, as well as the suggestions from expert validators, to refine the assessment instrument for evaluating high school students' critical thinking skills, which takes the form of an essay analysis test developed by the researcher.

b) Difficulty Level Analysis

The next step in the classical test analysis conducted by the researcher was the analysis of the difficulty level of the test items developed by the researcher. The research data showed that:

- **Pre-trial Phase:** The 10 essay analysis test items developed by the researcher had difficulty levels ranging between 0.35 and 0.82. This means that the test items developed had moderate to easy difficulty levels. According to Sugito et al. (2015: 83), a good test is one that is neither too easy nor too difficult. In other words, the test developed by the researcher is a good test.
- **Limited-scale Phase:** The 10 essay analysis test items developed by the researcher had difficulty levels ranging between 0.31 and 0.84. This means that the test items developed had a moderate difficulty level. According to Sugito et al. (2015: 83), a good test is one that is neither too easy nor too difficult. In other words, the test developed by the researcher is a good test.
- **Large-scale Phase:** The 10 essay analysis test items developed by the researcher had difficulty levels ranging between 0.43 and 0.93. This means that the test items developed had moderate to easy difficulty levels. According to Sugito et al. (2015: 83), a good test is one that is neither too easy nor too difficult. In other words, the test developed by the researcher is a good test.

c) Discriminatory Power Analysis

A good test is not only evaluated based on its validity and difficulty level but also on the discriminatory power of each item in measuring a student's ability to answer it. The discriminatory power analysis conducted on the essay analysis test items provided the following data:

Pre-trial	Limited scale	Large-scale stages
0,5	0,76	0,52
Good	Very Good	God

d) Reliability

The next analysis involved examining the reliability of the test items. According to Sugito et al. (2015: 101), a test is considered reliable if it consistently produces the same results when administered multiple times. This means that if a test is presented repeatedly, the results should be the same or very similar to previous outcomes. The research data showed that the reliability of the essay analysis test in the pre-trial phase was 0.71, in the limited-scale phase it was 0.72, and in the large-scale phase it was 0.74. In other words, overall, the test items developed by the researcher are fairly reliable as a tool for measuring students' critical thinking skills, as there was a significant improvement in the critical thinking instrument.

Data from the Problem-Solving Essay Test

In the development of the critical thinking skills test instrument, the researcher developed a problem-solving essay test consisting of 5 items. The data analysis was conducted using Microsoft Office Excel.

C) Difficulty Level Analysis

The next analysis involved the classical test analysis focusing on the difficulty level of each test item in measuring students' critical thinking skills. The data showed very clearly that the 5 test items developed by the researcher had difficulty levels ranging from 0.3 to 0.6 in the pre-trial phase and from 0.19 to 0.62. According to Sugito et al. (2015: 83), a good test is one that is neither too easy nor too difficult. This means that the test items developed by the researcher are suitable for use with the student sample.

Discriminatory Power Analysis

The data obtained from the classroom trials at the pre-trial, limited-scale, and large-scale stages in class XI MIA at SMA Percut Sei Tuan for the 2018/2019 academic year are as follows:

Table 9: Data from the Discriminatory Power Analysis of the Problem-Solving Test during the pre-trial, limited-scale, and large-scale stages.

Pre-trial	Limited scale	Large-scale stages
0,38	0,7	0,50
Sufficient	Very Good	Good

D) Reliability

The analysis conducted by the researcher was also supported by reliability data for the test items. In the critical thinking skills test, which was in the form of a problem-solving test, the pre-trial phase had a reliability score of 0.71, indicating high reliability. The limited-scale phase had a reliability score of 0.718, also indicating high reliability, and the large-scale phase had a reliability score of 0.79. This means that the problem-solving test is quite reliable for assessing students' critical thinking skills on the topic of sound waves. According to Surip (2014), critical thinking is the ability to solve problems, and it can be enhanced by improving critical thinking skills.

The testing conducted by the researcher on class XI MIA at SMA Negeri 2 Percut Sei Tuan for the 2018-2019 academic year showed that there were some valid and invalid data, reliable data, and measurement errors that were fairly reliable as evidence of classical test assessment. This data will support the accuracy of the researcher's analysis of critical thinking skills.

Research Discussion

The process of developing critical thinking skills instruments went through several stages of research and development, based on the stages used in Zahid's (2018) research. The first stage was the analysis phase, where data was collected on the types of instruments used by physics teachers in schools, the available facilities and infrastructure, and the learning process. Based on this data, there was potential to develop innovative instruments. The initial condition that required special attention was the type of physics instruments used in schools, which only measured memorization and understanding. This was not beneficial for students to practice critical thinking skills.

According to Ennis's indicators of critical thinking skills, the indicators that students predominantly answered were A1. Focusing on questions, A2. Asking and answering questions, and A3. Identifying assumptions. The indicators that were difficult for students to answer were A5. Defining terms and considering definitions, A7. Making and determining results of considerations, and A8. Observing and considering observation reports. Based on this, the researcher could identify the characteristics of the critical thinking skill indicators preferred by students to enhance their critical thinking abilities. For instance, the questions that were easy for them to answer were question 1, while the difficult ones were questions 4 and 6 for the essay analysis test. For the problem-solving test, the easy question was question 1, and the difficult ones were questions 2 and 3.

The analysis stage provided data on the types of physics assessment instruments, specifically for the sound wave topic in schools, which measure aspects of memorization and conceptual understanding. Lissa (2012) explained that based on Bloom's cognitive taxonomy, C1 (memorization) and C2 & C3 (understanding). Such conditions are certainly not ideal for practicing higher-order thinking skills in students. Developing learning instruments that focus on thinking skills is important due to advancements in science and technology. This aligns with Richmond's (2007) findings, which stated that good thinking skills could serve as a strong asset for students in Asia to face complex problems in the modern era. Such demands cannot easily be met without practice, as Yildrim & Ozkahraman (2011) emphasized that thinking skills can be developed through conditioning for thinking.

The thinking skills-oriented instruments were developed based on data from the initial analysis stage, which included the instruments used in schools, teaching systems, reviews of relevant studies, and reviews of government policies on national education orientation, as well as considering advancements in science and technology. The design of critical thinking skills instruments was adapted to Ennis's critical thinking indicators (1985). The developed instruments focused on critical thinking and problem-solving. The types of critical thinking skills assessment instruments developed were essay analysis tests and problem-solving tests.

The next stage was the development phase, where the first step involved expert validation. At this stage, the instruments designed were validated in terms of content by experts in thinking skills, educational research, and physics. The instruments validated included essay analysis tests, problem-solving tests, and questionnaires for student and teacher responses. Expert validation was conducted several times, as explained in the development stage. The general revisions involved writing style, indicator alignment, cognitive domain of the questions, accuracy of images, and terminology. The validation results were declared valid and usable in the field after revisions.

The type of validation measured was content validity, as critical thinking skills instruments should have good content validity before being used for measurement (Ennis & Weir, 1985; Doctor & Heller, 2009). Therefore, this stage is important because, before the instruments are tested, they must first be declared valid by experts to ensure that the obtained data is reliable. This development stage also included a pre-trial phase to measure the validity and reliability of the questions. The results were valid, with validity coefficients ranging from adequate to good. The reliability of the thinking questions was tested and declared reliable before use, in line with Suharsimi (2007).

The consistency of the questions in the essay analysis test and problem-solving test was somewhat lacking due to the type of questions focusing on thinking skills. Thinking skills instruments not only emphasize conceptual understanding but also synthesis, analysis, and evaluation aspects, leading to relatively low consistency (Carson, 2007; Doctor & Heller, 2009; Ennis & Weir, 1993). The trial timing was less supportive, as students were facing final school exams, leading to rushed answers. This outcome is consistent with Lissa's (2012) explanation that factors affecting instrument consistency include students' psychological conditions, unpleasant circumstances, cheating, questions being too brief, and questions being too easy or difficult. According to Mariani (2008), assessments relying on memorization questions tend to have high consistency or reliability, while thinking questions tend to have lower consistency.

This is because thinking questions rely on individual thinking abilities, which vary, whereas memorization questions have nearly the same standard answers for each individual. Since the critical thinking skills instrument was declared valid by experts, it was then tested in a limited-scale trial. During the limited-scale trial, reliability and student responses were measured. Developing thinking skills requires learning that activates students. The use of specific designs, such as conducting investigations, problem-solving, or questioning, can enhance thinking skills compared to conventional classrooms (Herman, 2007; Setiawan, 2008). In the limited-scale trial, the instrument was declared practical, with a high percentage of student questionnaires falling into the very high category. Practicality means that the instrument's use did not disrupt learning or interfere with the normal learning process, making the critical thinking skills instrument suitable for use.

In the large-scale trial, the improved instrument was tested in a class of approximately 30 students, and its reliability was calculated. Practicality was assessed from student response questionnaires, which were analyzed and showed a good category, with reliability reaching 0.8 in the high category. Although it falls into the good category, the instrument remains reliable, so the critical thinking skills assessment instrument can be used in the next stage. This stage is the implementation stage, where the effectiveness and practicality of the instrument are measured. The effectiveness indicator in the implementation phase is the positive impact of applying the critical thinking skills instrument on students' thinking abilities. Practicality has the same indicators as the limited-scale trial, but in the implementation phase, it is supplemented with teacher responses regarding the developed critical thinking skills instrument. Critical thinking skills are not an instant learning outcome that can be measured after two or three learning sessions and then declared good or not. According to Richmond (2007), developing critical thinking skills requires a process and practice that is not short.

At the beginning of the instrument testing, students were involved in contextual and problem-based activities. Students who were accustomed to passively receiving what was presented were gradually encouraged to contribute actively to the learning process. The instrument used to assess critical thinking and problem-solving skills should be based on basic knowledge. Carson (2007) explained that in problem-solving, the thinking process is more important than the knowledge itself, although basic knowledge is also an essential factor in solving a problem. Therefore, the development of critical thinking skills instruments was carried out without neglecting the concept.

Critical thinking skills instruments are part of the learning outcomes process that is not easy to improve. This is because not only critical thinking skills influence learning outcomes, but

also factors such as family conditions, economy, culture, and multicultural aspects (Lissa, 2012). Thus, to achieve good learning outcomes, factors other than critical thinking skills need to be strengthened. Students' thinking skills, whether in essay analysis tests or problem-solving tests, did not differ significantly in the implementation phase.

The critical thinking skills instrument was practical for use in learning. The development of the critical thinking skills assessment instrument has some limitations, including that the types of instruments developed only cover two skills, namely essay analysis tests and problem-solving tests. The second limitation is the use of indicators, but only those relevant to the research, and the developed instrument is still in the form of essays, which still give students the impression of being like a regular test.

Conclusions and Recommendations

Conclusions

Based on the discussion, the conclusions are as follows:

1. The development of critical thinking skills instruments for high school students was carried out through several steps: starting with defining the study, including preliminary analysis through field studies and preliminary studies. The product design phase involved preparing test blueprints, drafting questions, creating answer keys, and validating the design. The subsequent development phase included a small-scale pilot test, followed by a limited trial and a wide-scale implementation. The initial testing was conducted on a small scale, analyzed, and then implemented on a larger scale.
2. The relevant form of the critical thinking skills instrument for sound wave material consists of assessment instruments that measure critical thinking skills using indicators aligned with Bloom's taxonomy levels C4-C6. The assessment instruments included multiple-choice questions and essay tests; in this study, essay and problem-solving tests were used.
3. The characteristics of the critical thinking instruments, based on empirical data, showed that the validity of the essay analysis and problem-solving tests were rated as good. The reliability of both the essay analysis and problem-solving tests was also categorized as good. The questionnaire was deemed reliable with a Cronbach's Alpha > 0.70.

Recommendations

The development of the critical thinking skills assessment instrument has some limitations, including:

- The instrument developed only uses two types of skills: essay analysis and problem-solving tests.
- The use of critical thinking indicators and problem-solving does not cover all indicators but only those relevant to the research. The developed instrument remains in essay form, which may give students the impression of a typical test.

For future researchers conducting similar studies, it is recommended to address these limitations by:

- Expanding the types of critical thinking skills developed in the assessment instruments.
- Creating more engaging questions for students.
- Increasing the sample size for the implementation phase.

REFERENCES

Amalia, N., F., & Susilaningsih, E., (2014), Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Siswa Sma Pada Materi Asam Basa, *Jurnal Inovasi Pendidikan Kimia*, Vol 8 (2) : 1380-1389

Budiman, A., & Jailani., (2014), Pengembangan Instrumen Asesmen Higher Order Thinking Skill (Hots), *Jurnal Riset Pendidikan Matematatika*, Vol 1 (2) :1-13.

Carson, J. 2007. A. Problem With Problem Solving Teaching Thinking Without Teaching Knowledge. *The Mathematics Educators,* 17 (2):7-14.

Ennis, R., A., (1993), Critical Thinking Asessment, *Theory Into Practice*, Vol 32 (3) :179-186

Ennis, R., A., (1996), Critical Thinking Dispositions: Their Nature And Assesability, *Informal Logic*, Vol 18 (2) : 165-182.

Ennis, R., A., (1985), A Logical Basis For Measuring Critical Thinking Skill, Pacific Grove, Ca : Midwest Publication, I.

Ennis, R.H.& Weir, E.1985. *The Ennis Weir Critical Thinking Essay Test,* Pacific Grove, Ca: Midwest Publication, I.

Hadijah, A., (2016), Pengembangan Instrumen Tes Hasil Belajar Kognitif Mata, *Jurnal Pendidikan Fisika*, Vol 4 (1) : 1-5.

Herman, T.2007. Pembelajaran Berbasisi Masalah Untuk Meningkatkan Kemampuan Berpikir Matematis Tingkat Tinggi Siswa Smp. *Educationist*, 1(1):47-56.

Lissa . 2012.Pengembangan Instrument Penilaian Keterampilan Berpikir Tingkat Tinggi Materi System Respirasi Dan Ekskresi. *Jurnal Unnes Lik,* Vol 41 (1):27-32.

Jazuli, M., & Wardani, S., (2015), Pengembangan Alat Evaluasi Ipa Terpadu Topik Pembahasan Materi Berbasisi Kontekstual Untuk Menukur Kemampuan Berpikir Kritis Siswa, *Unnes Science Education Journal*, Vol 4 (2) : 1-7

Kopi, N., K.,E., Kendek, Y., & Ali, M., H., (2016), Penerapan Model Pembelajaran Penemuan Terbimbing Untuk Meningkatkan Hasil Belajar Fisika Pada Siswa Kelas 1-3.

Kartimi & Liliasari, (2012), Pengembangan Alat Ukur Berpikir Kritis Pada Konsep Termokimia Untuk Siswa Sma Peringkat Atas Dan Menengah, *Jurnal Pendidikan Ipa Indonesia*, Vol 1 :21-26

Mardapi, D., (2008), *Tehnik Penyusunan Instrument Tes Dan Non Tes,* Mitra Cindekia, Yogyakarta.

Prasasti, Y., R., Suyono, & Basuki, (2012), Pengembangan Instrument Asesment Berfikir Kritis Melalui Membaca Untuk Siswa Sd/Mi, *Jurnal Universitas Negeri Malang*, Vol 48 (2) :1-12.

Purwanto, M., N., (2013), Psikologi Pendidikan, Pt Remaja Rosdakarya, Bandung.

Redhana, I., W., & Liliasari, (2008), Program Pembelajaran Keterampilan Berpikir Pada Topik Laju Reaksi Untuk, *Jurnal Forum Kependidikan,*Vol 27 (2) : 103-112.

Reta, I., K., (2012), Pengaruh Model Pembelajran Berbasis Masalah Terhadap Keterampilan Berpikir Kritis Ditinjau Dari Gaya Kognitif Siswa, *Jurnal Pendidikan,* Vol 26 (1) : 1-16.

Rusyna, A., (2014), *Keterampilan Berfikir,*Ombak, Yogyakarta.

Richmond , J.E.D.2007.Bringing Critical Thinking To The Education Of Developing Country Professionals. *International Education Journal*, 8 (1):1-29.

Salim & Syahrum, (2012), *Metodologi Penelitian,* Citapustaka Media, Bandung.

Sudjana, N., & Ibrahim, (2014), *Penelitian Dan Penilaian Pendidikan,* Percetakan Sinar Baru Algensindo Offset, Bandung.

Sugiyono, (2017), *Statistika Untuk Penelitian,* Alfabeta, Bandung.

Sugiyono, (2010), *Metode Penelitian: Kuantitatif, Kualitatif Dan R&D*, Alfabeta, Bandung.

Sukma Pradana, S.,D., Parno & Handayanto, S., K., (2017), Pengembangan Tes Kemampuan Berfikir Kritis Pada Materi Optik Geometri Untuk Mahasisiwa Fisika, *Jurnal Penelitian Dan Evaluasi*, Vol:21 (1)1-14.

Surip, M., (2014), *Berfikir Kritis,* Halaman Moeka Publishing, Jakarta.

Suharsimi. 2007. Dasar-Dasar Evaluasi Pendidikan (Edisi Revisi). Bumi Aksara.

Syafi'i, A., Handayani, L., & Khanafiyah, S., (2014), Penerapan Question Based Discovery Learning Pada Kegiatan Laboratorium Fisika Untuk Meningkatkan Keterampiulan Proses Sains, *Unnes Physics Education Journal*, Vol 3 (2) : 1-8.

Sari, W., Jufrida, & Pathoni, H., (2017), Pengembangan Modul Elektronik Berbasis 3D Pagaeflip Professional Pada Materi Konsep Dasar Fisika Inti Dan Struktur Inti Mata Kuliah Fisika Atom Dan Inti*, Jurnal Edufisika,* Vol:2 (1) 1-13.

Tanjung, R., (2017), *Evaluasi Hasil Belajar,* Unimed Press, Medan.

Trianto, (2018), *Mendesain Model Pembelajaran Inovatif-Progresif,* Kharisma Putra Utama, Jakarta.

Yildrim, B. & Ozkahraman, S. 2011. Critical Thinking In Nursing Process And Education . International Journal Of Humanities And Social Science, 1 (13):257-262.