International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

Development Of Critical Thinking Skills Instruments In Elementary School Mathematics Problem Solving

Rosdiana
Universitas Negeri Makassar, Indonesia
Syahrul
Universitas Negeri Makassar, Indonesia
Baso Intang Sappaile
Universitas Negeri Makassar, Indonesia

Corresponding Author: rosdiana655@gmail.com; syahrul@unm.ac.id; baso.sappaile@unm.ac.id

Abstract

This study aims to develop an assessment instrument for measuring elementary school students' critical thinking skills in mathematics problem solving that is valid, reliable, and practical for teachers to use. The study was motivated by the need for an assessment instrument that does not merely evaluate students' final answers, but also reveals their critical thinking processes in understanding problems, planning strategies, carrying out solutions, and reviewing results. The instrument was developed based on critical thinking indicators, including interpretation, analysis, inference, evaluation, explanation, and self-regulation, integrated with Polya's stages of problem solving. This research employed a research and development design using the ADDIE model, consisting of analysis, design, development, implementation, and evaluation stages. The trial subjects were fifth-grade elementary school students in the Selayar Islands Regency. The product developed was an open-ended essay test equipped with a test blueprint, instructions, and an analytical scoring rubric. Data were collected through observation, interviews, expert validation, instrument trials, and practicality questionnaires. The data were analyzed using content validity, construct validity through Confirmatory Factor Analysis (CFA), Cronbach's Alpha reliability, item difficulty analysis, item discrimination analysis, and practicality analysis. The results showed that the developed instrument had very good content validity, with S-CVI values of 0.875 for the material aspect, 0.925 for the construct aspect, 0.917 for the language aspect, and 0.958 for the scoring rubric. The construct validity also demonstrated a good model fit, with CFI = 0.979, TLI = 0.971, and RMSEA = 0.074, while all factor loadings were above 0.80. The reliability of the

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

instrument was categorized as very high, with a Cronbach's Alpha value of 0.950 for the 10 essay items. In addition, the instrument was considered highly practical based on expert assessment at 88.13%, teacher responses at 85.94%, implementation observation at 85.16%, and an overall average of 86.41%. Therefore, the developed instrument for assessing critical thinking skills in elementary school mathematics problem solving is valid, reliable, and practical for use as a formative, diagnostic, and limited summative assessment tool in elementary mathematics learning.

Keywords: Assessment Instrument, Critical Thinking, Mathematics Problem Solving, Elementary School, ADDIE.

INTRODUCTION

The development of education in the twenty-first century demands students not only master basic knowledge, but also possess higher-order thinking skills (HOTS), particularly critical thinking skills. In the era of rapid information growth and complex social changes, students are expected to analyze information critically, evaluate arguments logically, and solve problems systematically. Critical thinking has therefore become one of the essential competencies that must be developed from the elementary school level because it forms the foundation of students' intellectual and problem-solving abilities. According to Facione (2015), critical thinking is a reflective and rational thinking process used to make reasonable judgments and decisions. In the educational context, critical thinking enables students to interpret information, analyze relationships among concepts, draw logical conclusions, and evaluate evidence critically. Consequently, learning activities should not merely focus on memorization and procedural knowledge but also encourage students to engage in meaningful reasoning processes.

Mathematics is considered one of the most strategic subjects for fostering students' critical thinking skills because it emphasizes logical reasoning, analytical thinking, and problem solving. Mathematics learning encourages students to identify patterns, formulate strategies, and evaluate solutions systematically. Organisation for Economic Co-operation and Development stated that mathematical literacy involves the ability to formulate, employ, and interpret mathematics in various contexts, including reasoning mathematically and using mathematical concepts to solve real-life problems (OECD, 2019). Thus, mathematics education should not only aim at obtaining correct answers but also facilitate students in developing reasoning and reflective thinking skills through problem-solving activities. Polya (1973) emphasized that mathematical

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

problem solving consists of four important stages: understanding the problem, devising a plan, carrying out the plan, and looking back at the solution. These stages indicate that problem solving inherently requires critical thinking processes.

However, various international and national assessment results indicate that Indonesian students' critical thinking and mathematical problem-solving skills remain relatively low. The results of the Programme for International Student Assessment (PISA) and the Asesmen Kompetensi Minimum (AKM) revealed that many Indonesian elementary students still struggle to solve mathematical problems that require reasoning, analysis, and evaluation (Kemendikbudristek, 2023; OECD, 2019). This condition reflects that mathematics learning in elementary schools has not fully facilitated the development of higher-order thinking skills. Based on preliminary observations and interviews with elementary school teachers in Selayar Islands Regency, most mathematics assessments still rely on conventional tests emphasizing correct answers rather than students' reasoning processes. Teachers frequently use multiple-choice and short-answer questions that mainly assess lower cognitive levels (C1–C3), while assessment instruments specifically designed to measure critical thinking skills are rarely utilized.

The lack of appropriate assessment instruments has become one of the major challenges in evaluating students' critical thinking abilities in mathematics problem solving. Existing assessments generally focus on procedural competence and final results rather than students' analytical and reflective thinking processes. According to Ennis (2011), critical thinking involves several important dimensions, including interpretation, analysis, inference, evaluation, and explanation. Therefore, assessment instruments intended to measure critical thinking should capture students' reasoning processes comprehensively. Nevertheless, previous studies revealed that most critical thinking instruments developed for elementary education still emphasize content validity and have not comprehensively tested construct validity and reliability using advanced psychometric analyses such as Confirmatory Factor Analysis (CFA). Dewi, Sariyasa, and Putrayasa (2020) developed a critical thinking instrument with high content validity, yet the instrument lacked deeper construct validation. Similarly, Yudha (2019) applied Rasch modeling in developing critical thinking assessments, but the study focused on vocational high school students rather than elementary school learners whose cognitive characteristics differ significantly.

Another important issue is the limited integration between critical thinking indicators and mathematical problem-solving processes in existing assessment instruments. Many instruments use general indicators such as analysis and evaluation without specifically relating them to Polya's

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

problem-solving stages. As a result, teachers experience difficulties in designing assessment tasks capable of measuring students' critical thinking throughout the entire mathematical problem-solving process. Furthermore, the implementation of the Kurikulum Merdeka emphasizes authentic, formative, and diagnostic assessments that support the development of higher-order thinking skills. However, many teachers still lack practical guidelines and comprehensive scoring rubrics to assess students' critical thinking consistently and objectively.

Based on these theoretical, methodological, and contextual gaps, this study focuses on developing an assessment instrument for measuring elementary school students' critical thinking skills in mathematics problem solving. The instrument integrates critical thinking indicators proposed by Ennis (2011) and Facione (2015) with Polya's (1973) mathematical problem-solving stages. Moreover, the instrument is developed using the ADDIE model, consisting of analysis, design, development, implementation, and evaluation stages. Through this systematic process, the study aims to produce an instrument that is valid, reliable, and practical for classroom implementation. The developed instrument is expected to support teachers in conducting authentic assessments, identifying students' critical thinking patterns, and improving the quality of mathematics learning in elementary schools. Ultimately, this research contributes theoretically to the field of educational evaluation and practically to the implementation of critical-thinking-based mathematics assessment aligned with the principles of the Kurikulum Merdeka.

Critical Thinking Theory

Critical thinking is one of the most essential competencies in twenty-first-century education because it enables learners to analyze information, evaluate evidence, and make logical decisions in solving problems. In the context of elementary education, critical thinking is not limited to students' ability to provide correct answers, but also includes their ability to understand problems, analyze relationships among concepts, formulate arguments, and reflect on their reasoning processes. According to Facione (2015), critical thinking is a purposeful, self-regulatory judgment that results in interpretation, analysis, evaluation, inference, and explanation. This theory emphasizes that critical thinking is an active cognitive process involving reflective reasoning and logical judgment. Facione further explained that individuals with critical thinking skills are able to examine assumptions, evaluate arguments, and draw rational conclusions based on available evidence. Therefore, critical thinking is considered a higher-order cognitive process that requires students to engage deeply with information rather than simply memorizing facts.

Similarly, Ennis (2011) defined critical thinking as "reasonable reflective thinking focused

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

on deciding what to believe or do." This definition highlights two important dimensions of critical thinking: reasonableness and reflectiveness. Reasonableness refers to the ability to make decisions based on logical evidence, while reflectiveness involves careful consideration of assumptions, arguments, and possible solutions before making judgments. Ennis proposed several indicators of critical thinking, including interpretation, analysis, inference, evaluation, explanation, and self-regulation. These indicators serve as important foundations in developing educational assessments that aim to measure students' thinking processes systematically. In mathematics education, these dimensions are highly relevant because students are frequently required to interpret mathematical information, analyze relationships among quantities, infer solutions, and evaluate the validity of their answers.

The importance of critical thinking in mathematics learning is also strongly associated with the development of Higher Order Thinking Skills (HOTS). HOTS involve students' abilities to analyze, evaluate, and create knowledge rather than merely remembering or understanding information. According to the revised Bloom's Taxonomy, higher-order cognitive processes include analyzing (C4), evaluating (C5), and creating (C6). These cognitive levels correspond closely with critical thinking dimensions because they require students to engage in reasoning, reflection, and problem solving. In the context of elementary mathematics learning, students are expected not only to solve routine problems procedurally but also to explain their reasoning, justify their solutions, and evaluate alternative strategies.

The relevance of critical thinking theory in mathematics education is reinforced by the demands of the Kurikulum Merdeka, which emphasizes competency-based learning and authentic assessment. The curriculum encourages teachers to facilitate students' critical and creative thinking through contextual learning activities and formative assessments. In this context, assessment instruments should not merely measure final outcomes but should also capture students' reasoning processes during problem solving. Therefore, the integration of critical thinking theory into mathematics assessment becomes highly important to ensure that students' higher-order cognitive skills are adequately evaluated.

Several previous studies have highlighted the significance of critical thinking assessment in mathematics learning. Dewi, Sariyasa, and Putrayasa (2020) emphasized that critical thinking assessments can improve students' engagement in analytical reasoning and reflective problem solving. However, they also noted that many existing instruments still lack comprehensive integration between critical thinking indicators and mathematical problem-solving processes. Likewise, Yudha (2019) found that authentic assessment instruments based on critical thinking

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

theory could effectively measure students' reasoning skills, although the instruments were primarily developed for secondary education. These findings suggest the necessity of developing contextually appropriate instruments specifically designed for elementary school students.

From a theoretical perspective, critical thinking theory provides a strong conceptual foundation for developing assessment instruments capable of measuring students' reasoning processes in mathematics problem solving. The indicators proposed by Ennis (2011) and Facione (2015) enable educators to design assessment tasks that evaluate students' abilities to interpret problems, analyze information, infer logical conclusions, evaluate arguments, and explain reasoning systematically. Consequently, critical thinking theory becomes a crucial framework in understanding how students process mathematical information and solve complex problems reflectively and logically.

Mathematical Problem-Solving Theory

Mathematical problem solving is considered one of the central goals of mathematics education because it promotes reasoning, logical thinking, and analytical skills. Problem solving enables students to apply mathematical concepts in meaningful contexts while developing their ability to think critically and systematically. One of the most influential theories of mathematical problem solving was proposed by George Polya in his seminal work How to Solve It (1973). Polya explained that effective problem solving involves four interconnected stages: understanding the problem, devising a plan, carrying out the plan, and looking back. These stages provide a structured framework that guides students through the cognitive processes required to solve mathematical problems effectively.

The first stage, understanding the problem, requires students to identify known information, determine what is being asked, and interpret the relationships among mathematical elements within the problem. At this stage, students engage in interpretation and comprehension processes that are closely related to critical thinking skills. Without a clear understanding of the problem, students may apply incorrect procedures or fail to identify relevant information. Therefore, comprehension becomes the foundation for successful problem solving.

The second stage, devising a plan, involves selecting appropriate strategies or procedures to solve the problem. Students must analyze possible methods, compare alternative approaches, and determine the most effective strategy based on the information available. This stage reflects analytical and inferential thinking processes because students must reason logically before implementing solutions. Common strategies suggested by Polya include drawing diagrams,

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

identifying patterns, simplifying problems, and working backward.

The third stage, carrying out the plan, focuses on implementing the chosen strategy systematically and accurately. At this stage, procedural knowledge and logical reasoning work together as students execute mathematical operations and verify intermediate results. Students must remain attentive to potential errors and inconsistencies during the process. This stage demonstrates that mathematical problem solving is not purely mechanical but requires continuous monitoring and reflective thinking.

The final stage, looking back, emphasizes reflection and evaluation of the solution obtained. Students are encouraged to review their answers, verify the correctness of procedures, and consider whether alternative methods might yield better solutions. According to Polya (1973), reflection is essential because it develops students' awareness of their own thinking processes and strengthens their ability to evaluate mathematical reasoning critically. This stage aligns closely with the evaluation and self-regulation dimensions of critical thinking proposed by Facione (2015).

The integration of Polya's problem-solving stages into mathematics education has significant implications for assessment practices. Traditional assessments that focus solely on final answers often fail to capture students' reasoning processes during problem solving. Consequently, many educators advocate for authentic assessments that evaluate students' thinking at every stage of the problem-solving process. In elementary mathematics education, contextual and open-ended essay questions are considered particularly effective because they allow students to demonstrate their reasoning, explain strategies, and reflect on solutions comprehensively.

Research findings also support the importance of problem-solving-based assessment. Ndiung and Jediut (2020) found that mathematics learning emphasizing HOTS-based problem solving significantly improved students' reasoning abilities. Similarly, OECD (2019) highlighted that mathematical literacy involves students' abilities to formulate, employ, and interpret mathematics in various contexts. This perspective reinforces the view that mathematics education should emphasize reasoning and problem solving rather than rote memorization and procedural calculation alone.

In relation to this study, Polya's theory provides the operational framework for designing assessment instruments capable of measuring critical thinking within mathematical problem solving. Each problem-solving stage corresponds to specific critical thinking indicators, enabling educators to evaluate students' reasoning systematically. Understanding problems reflects interpretation skills; devising plans reflects analytical and inferential thinking; carrying out plans demonstrates logical reasoning; and looking back reflects evaluative and reflective thinking.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Therefore, Polya's theory serves as an essential theoretical foundation for integrating critical thinking into mathematics assessment.

Educational Assessment Theory

Educational assessment is a systematic process used to collect, interpret, and evaluate information regarding students' learning outcomes and competencies. In contemporary education, assessment is not merely intended to measure academic achievement but also to support learning processes, diagnose difficulties, and improve instructional quality. According to educational evaluation theory, effective assessment should be valid, reliable, practical, and capable of measuring intended competencies comprehensively. In the context of mathematics education, assessment should evaluate not only procedural knowledge but also students' reasoning, problem-solving abilities, and higher-order thinking skills.

Authentic assessment has become increasingly important in modern educational practices because it focuses on real-world applications and meaningful learning experiences. Authentic assessment requires students to demonstrate their understanding through performance tasks, problem solving, and reasoning activities rather than selecting answers from predetermined options. In mathematics learning, authentic assessments often involve open-ended problems that encourage students to explain their thinking processes and justify their solutions logically. This approach aligns with the principles of the Kurikulum Merdeka, which emphasizes formative and diagnostic assessment practices aimed at supporting competency development.

Assessment validity is one of the most important concepts in educational measurement. Validity refers to the extent to which an instrument measures what it is intended to measure. According to the thesis, content validity can be evaluated using Content Validity Ratio (CVR) and Content Validity Index (CVI), while construct validity can be examined through Confirmatory Factor Analysis (CFA). Construct validity is particularly important because it ensures that assessment items accurately represent the theoretical dimensions of critical thinking being measured. In this study, CFA is used to confirm whether the assessment items correspond to the critical thinking indicators proposed by Ennis (2011) and Facione (2015).

Reliability is another essential characteristic of high-quality assessment instruments. Reliability refers to the consistency and stability of assessment results across different contexts and administrations. Cronbach's Alpha is commonly used to measure internal consistency reliability. A high reliability coefficient indicates that assessment items consistently measure the same construct. In educational assessment, reliable instruments are necessary to ensure fair and

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

objective evaluation of students' abilities.

In addition to validity and reliability, item analysis also plays an important role in educational assessment development. Item difficulty and discrimination indices help determine whether assessment items are appropriately challenging and capable of differentiating between students with high and low abilities. According to Classical Test Theory, moderate difficulty levels and strong discrimination indices contribute to effective assessment instruments. These psychometric properties ensure that assessment tasks accurately reflect students' critical thinking competencies.

The development of assessment instruments also requires systematic procedures to ensure product quality. In this study, the ADDIE model is used as the framework for instrument development. The ADDIE model consists of Analysis, Design, Development, Implementation, and Evaluation stages. During the analysis stage, researchers identify students' needs, curriculum demands, and classroom conditions. In the design stage, assessment blueprints and scoring rubrics are prepared. The development stage focuses on constructing and validating assessment items, while implementation involves limited field testing. Finally, evaluation is conducted to revise and improve the instrument based on empirical findings.

Educational assessment theory therefore provides the methodological and psychometric foundation for developing valid and reliable instruments capable of measuring critical thinking skills in mathematics problem solving. Through systematic assessment procedures, educators can obtain comprehensive information about students' reasoning abilities and use the results to improve instructional practices. Consequently, assessment becomes not only a tool for measuring learning outcomes but also a means of supporting meaningful learning and developing students' higher-order thinking skills..

DISCUSSION

The Needs And Characteristics Of An Assessment Instrument For Critical Thinking Skills In Mathematical Problem Solving At The Elementary School Level

The first research problem focuses on identifying the needs and characteristics of an assessment instrument designed to measure elementary school students' critical thinking skills in mathematical problem solving. This issue emerges from the growing demand for higher-order thinking skills (HOTS) in twenty-first-century education, particularly in mathematics learning. The thesis explains that current mathematics assessments in elementary schools are still dominated by conventional tests emphasizing memorization and procedural accuracy rather than analytical

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

reasoning and reflective thinking. As stated in the study, most teachers continue to use multiple-choice and short-answer tests that mainly measure lower-order cognitive skills (C1–C3), which are insufficient for evaluating students' critical thinking abilities. This condition reflects a significant gap between curriculum expectations and actual classroom assessment practices.

The analysis of the first problem indicates that the need for a critical-thinking-based assessment instrument is strongly related to the implementation of the Kurikulum Merdeka, which emphasizes competency-based learning and authentic assessment. The curriculum encourages teachers to develop students' reasoning, analytical, and evaluative skills through contextual learning activities. However, the thesis reveals that many teachers still encounter difficulties in designing assessment tasks capable of measuring students' critical thinking systematically. Existing assessments generally focus on final answers rather than students' reasoning processes in solving mathematical problems. This finding is consistent with Facione's (2015) theory, which defines critical thinking as a process involving interpretation, analysis, evaluation, inference, explanation, and self-regulation. Therefore, an effective assessment instrument should measure not only cognitive outcomes but also the thinking processes underlying students' responses.

The thesis further analyzes that the characteristics of the desired instrument should align with both critical thinking theory and mathematical problem-solving theory. In this context, the study integrates the critical thinking indicators proposed by Ennis (2011) and Facione (2015) with Polya's (1973) stages of problem solving, namely understanding the problem, devising a plan, carrying out the plan, and looking back. This integration is important because mathematical problem solving inherently requires students to interpret information, analyze relationships among concepts, infer logical conclusions, and evaluate solutions critically. Consequently, the instrument must contain contextual and open-ended problems that encourage students to explain their reasoning processes comprehensively.

Another important aspect identified in the analysis is the need for assessment instruments that are psychometrically sound. Previous studies cited in the thesis, such as Dewi, Sariyasa, and Putrayasa (2020) and Yudha (2019), demonstrated that many critical thinking instruments still focus primarily on content validity without comprehensive construct validation and reliability testing. The thesis argues that a high-quality assessment instrument should fulfill criteria of validity, reliability, practicality, item difficulty, and discrimination power. Therefore, the instrument developed in this study is expected to undergo rigorous psychometric testing, including Content Validity Index (CVI), Confirmatory Factor Analysis (CFA), and Cronbach's Alpha

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

reliability analysis.

The findings from the needs analysis also reveal that teachers require practical guidelines and analytic scoring rubrics to assess students' critical thinking objectively and consistently. The absence of standardized rubrics often causes subjective judgments and inconsistent scoring practices among teachers. Therefore, the proposed instrument is designed not only as a test but also as a comprehensive assessment package consisting of a test blueprint, contextual essay questions, scoring rubrics, and implementation guidelines. Such characteristics are intended to support authentic and formative assessment practices in elementary mathematics learning.

Overall, the analysis of the first research problem demonstrates that the development of a critical-thinking-based mathematics assessment instrument is urgently needed to address the limitations of current assessment practices. The instrument must reflect critical thinking dimensions, integrate mathematical problem-solving processes, fulfill psychometric standards, and remain practical for classroom implementation. By fulfilling these requirements, the instrument is expected to support teachers in evaluating students' higher-order thinking skills more comprehensively and meaningfully, while also contributing to the improvement of mathematics learning quality in elementary schools.

The Process Of Developing An Assessment Instrument For Critical Thinking Skills In Mathematical Problem Solving For Elementary School Students Using The Addie Development Model

The second research problem focuses on the process of developing an assessment instrument for measuring critical thinking skills in mathematical problem solving among elementary school students using the ADDIE development model. This issue is highly significant because the quality of educational assessment instruments greatly influences the accuracy of measuring students' higher-order thinking skills. In mathematics education, assessment should not merely evaluate students' final answers, but also examine the reasoning processes used in understanding problems, planning strategies, implementing solutions, and reflecting on outcomes. Therefore, the development of a systematic and theoretically grounded instrument is necessary to ensure that critical thinking skills can be measured comprehensively and objectively.

The study adopts the ADDIE model, consisting of Analysis, Design, Development, Implementation, and Evaluation stages, as the primary framework for instrument development. The use of the ADDIE model reflects a systematic approach that allows researchers to produce assessment instruments that are valid, reliable, and practical. According to the thesis, the analysis

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

stage involves identifying curriculum requirements, students' cognitive characteristics, teachers' assessment needs, and the challenges encountered in mathematics learning. This stage is essential because the success of an assessment instrument depends on its alignment with the actual learning context and educational objectives. The study emphasizes that elementary school mathematics assessments are still largely dominated by conventional tests that focus on lower-order cognitive skills rather than critical thinking processes. Consequently, the analysis stage provides the empirical and theoretical foundation for developing a more authentic assessment instrument.

The design stage involves constructing indicators, test blueprints, essay questions, and scoring rubrics based on critical thinking theory and mathematical problem-solving theory. The thesis explains that the instrument integrates critical thinking indicators proposed by Ennis (2011) and Facione (2015), including interpretation, analysis, inference, evaluation, explanation, and self-regulation, with Polya's (1973) mathematical problem-solving stages. This integration is analytically important because it ensures that each assessment item measures not only procedural competence but also students' reasoning processes during mathematical problem solving. Through contextual and open-ended essay questions, students are encouraged to explain their thinking, justify their reasoning, and evaluate their solutions critically. This approach is consistent with authentic assessment principles promoted in the Kurikulum Merdeka, which emphasizes formative and competency-based assessment.

The development stage focuses on constructing and validating the assessment instrument. Expert validation is conducted to examine the content, language, and construct appropriateness of the instrument. The thesis highlights that content validity is measured using Content Validity Index (CVI), while construct validity is tested using Confirmatory Factor Analysis (CFA). This analytical procedure demonstrates methodological rigor because it goes beyond traditional content validation by examining the empirical relationship between assessment items and theoretical constructs. In addition, reliability testing using Cronbach's Alpha ensures that the instrument consistently measures critical thinking skills. The use of psychometric analyses such as CFA and reliability testing strengthens the scientific quality of the developed instrument and distinguishes this study from many previous studies that focused only on basic validity measures.

The implementation stage involves limited field testing with fifth-grade elementary school students in Selayar Islands Regency. This stage allows researchers to collect empirical data regarding item difficulty, discrimination index, practicality, and students' responses to the assessment tasks. The implementation process is particularly important because it evaluates

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

whether the instrument functions effectively in real classroom conditions. Teachers' responses and observations also contribute to determining the practicality of the instrument in supporting authentic mathematics assessment.

Finally, the evaluation stage focuses on revising and improving the instrument based on validation and field-testing results. This stage reflects the iterative nature of educational product development because assessment instruments require continuous refinement to achieve optimal quality. Overall, the ADDIE-based development process demonstrates a comprehensive and systematic approach to creating a critical-thinking-based mathematics assessment instrument. The study contributes not only to educational assessment theory but also to practical classroom assessment by providing teachers with a valid, reliable, and contextually relevant tool for evaluating students' critical thinking skills in mathematical problem solving.

The Results Of The Content Validity And Construct Validity Of The Developed Assessment Instrument For Critical Thinking Skills In Elementary School Mathematical Problem Solving

The third research problem in this study focuses on examining the content validity and construct validity of the developed instrument for assessing critical thinking skills in elementary school mathematics problem solving. This aspect is highly significant because validity determines whether an instrument accurately measures the intended construct. In educational assessment, validity is considered the foundation of a high-quality instrument because invalid instruments may lead to inaccurate interpretations of students' abilities. The thesis emphasizes that the development of the instrument was grounded in critical thinking theories proposed by Ennis (2011) and Facione (2015), integrated with Polya's (1973) mathematical problem-solving stages. Therefore, the validation process aimed to ensure that every test item genuinely reflected the dimensions of critical thinking and mathematical reasoning required in the study.

The analysis of content validity in the study demonstrates that the instrument possesses a very high level of expert agreement regarding the relevance and representativeness of the assessment items. According to the findings presented in the thesis, the instrument achieved S-CVI values of 0.875 for the material aspect, 0.925 for the construct aspect, 0.917 for the language aspect, and 0.958 for the scoring rubric. These results indicate that the developed instrument meets the acceptable standards of content validity because the values exceed the minimum threshold commonly recommended in educational measurement literature. Content validity is particularly important because it confirms that the assessment items are aligned with curriculum objectives, critical thinking indicators, and students' cognitive characteristics. The strong content

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

validity found in this study suggests that the instrument successfully represents the theoretical dimensions of critical thinking, including interpretation, analysis, inference, evaluation, explanation, and self-regulation, as proposed by Facione (2015).

In addition to content validity, construct validity was examined using Confirmatory Factor Analysis (CFA), which is regarded as one of the most robust statistical techniques for validating psychological and educational constructs. Construct validity refers to the extent to which the instrument accurately measures the theoretical construct it is intended to assess. In this study, CFA was used to verify whether the empirical data supported the theoretical framework integrating critical thinking dimensions with Polya's problem-solving stages. The results showed excellent model fit indices, with Comparative Fit Index (CFI) = 0.979, Tucker Lewis Index (TLI) = 0.971, and Root Mean Square Error of Approximation (RMSEA) = 0.074. These indices indicate that the measurement model fits the empirical data very well because CFI and TLI values above 0.90 and RMSEA values below 0.08 are generally considered indicators of good model fit in educational research.

Furthermore, the thesis reported that all factor loadings exceeded 0.80, demonstrating that every item strongly contributed to the measurement of critical thinking skills in mathematical problem solving. High factor loadings indicate that the assessment items are highly representative of the latent construct being measured. This finding is important because it confirms that the developed instrument is not merely a collection of mathematics questions but a theoretically grounded assessment tool capable of evaluating students' higher-order thinking processes systematically. The integration of critical thinking dimensions and Polya's stages appears to have strengthened the conceptual coherence of the instrument, making it more suitable for authentic assessment in elementary mathematics education.

The use of CFA in this study also distinguishes the research from many previous studies that relied solely on content validity without examining deeper psychometric properties. Earlier studies discussed in the thesis, such as those by Dewi et al. (2020) and Yudha (2019), primarily focused on content validity and reliability but lacked comprehensive construct validation. By incorporating CFA, this study provides stronger empirical evidence regarding the structural validity of the instrument. Consequently, the developed instrument not only fulfills theoretical expectations but also demonstrates empirical consistency between conceptual dimensions and observed student responses.

Overall, the findings related to the third research problem indicate that the developed instrument possesses strong psychometric quality in terms of both content validity and construct

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

validity. These results imply that the instrument is suitable for assessing elementary students' critical thinking skills in mathematical problem solving within the framework of authentic and competency-based assessment promoted by the Kurikulum Merdeka. The rigorous validation procedures employed in this study contribute significantly to the field of educational evaluation by providing a valid and theoretically sound instrument capable of supporting formative, diagnostic, and limited summative assessment practices in elementary mathematics learning.

The Levels Of Reliability, Item Difficulty, And Discrimination Index Of The Developed Assessment Instrument For Critical Thinking Skills In Elementary School Mathematical Problem Solving

The fourth research problem in this study examines the reliability, item difficulty level, and discrimination index of the developed instrument for assessing critical thinking skills in elementary school mathematics problem solving. These three psychometric aspects are crucial in determining the quality and effectiveness of an educational assessment instrument. Reliability ensures that the instrument consistently measures students' abilities across different conditions, while item difficulty and discrimination indices provide information about the appropriateness and effectiveness of each test item in differentiating students according to their levels of ability. In the context of critical thinking assessment, these components become particularly important because the instrument is intended not only to evaluate procedural mathematical competence but also to measure higher-order cognitive processes such as analysis, evaluation, inference, and reflective reasoning.

The study found that the developed instrument demonstrated a very high level of reliability, indicated by a Cronbach's Alpha coefficient of 0.950 for the ten essay items included in the test. According to educational measurement standards, a reliability coefficient above 0.80 is categorized as very high and indicates strong internal consistency among test items. This finding suggests that the instrument consistently measures the same construct, namely critical thinking skills in mathematical problem solving. The high reliability coefficient also implies that students' scores are stable and dependable, reducing the likelihood of measurement error. In assessment theory, reliability is essential because inconsistent instruments may produce inaccurate interpretations of students' competencies. Therefore, the Cronbach's Alpha result obtained in this study confirms that the developed instrument can be trusted as an effective tool for evaluating elementary students' critical thinking skills in mathematics learning.

The high reliability of the instrument can be attributed to several factors discussed in the thesis. First, the instrument was systematically developed based on strong theoretical

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

foundations, integrating critical thinking indicators proposed by Ennis (2011) and Facione (2015) with Polya's (1973) mathematical problem-solving stages. This integration ensured conceptual consistency across the test items, enabling the instrument to measure related dimensions of critical thinking comprehensively. Second, the use of analytical scoring rubrics helped standardize the evaluation process, minimizing subjectivity in scoring students' responses. Because essay tests are often vulnerable to scorer bias, the provision of detailed scoring criteria significantly contributed to improving scoring consistency and overall reliability.

Another important aspect analyzed in the study is the level of item difficulty. Item difficulty refers to the extent to which test items are easy or difficult for students to answer correctly. According to Classical Test Theory, effective assessment instruments should contain items with moderate difficulty levels to ensure that students with varying abilities can engage meaningfully with the tasks. The thesis explains that the developed instrument achieved item difficulty levels within the acceptable moderate range, specifically between 0.30 and 0.70. This finding indicates that the test items were neither too easy nor excessively difficult for elementary school students. Moderate difficulty levels are important because overly easy questions fail to challenge students' critical thinking processes, while excessively difficult questions may discourage students and fail to provide accurate information about their abilities.

The contextual nature of the test items likely contributed to the appropriate difficulty levels observed in the study. The instrument used open-ended essay questions based on real-life mathematical situations familiar to elementary students, such as buying and selling activities, measurements, and fraction comparisons. These contextualized tasks allowed students to connect mathematical concepts with everyday experiences, making the problems more meaningful and accessible while still requiring analytical reasoning and critical thinking. This approach aligns with the principles of authentic assessment emphasized in the Kurikulum Merdeka, which encourages meaningful and competency-based learning experiences.

In addition to item difficulty, the discrimination index was analyzed to determine how effectively each item differentiated between high-performing and low-performing students. A good discrimination index indicates that students with stronger critical thinking abilities perform significantly better on the item than students with weaker abilities. According to the thesis, the developed instrument achieved discrimination indices above the minimum acceptable level of 0.30, indicating that the items effectively distinguished students according to their levels of critical thinking competence. This result is particularly important because one of the primary goals of assessment is to identify variations in students' abilities accurately. Strong discrimination indices

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

demonstrate that the instrument is sensitive enough to detect differences in students' reasoning and problem-solving skills.

The findings regarding reliability, item difficulty, and discrimination index collectively demonstrate that the developed instrument possesses strong psychometric quality. Unlike many previous studies that focused primarily on content validity, this research comprehensively evaluated the empirical characteristics of the assessment instrument using systematic statistical analyses. As discussed in the thesis, earlier studies such as Dewi et al. (2020) and Marselina et al. (2021) often limited their analyses to basic validity and reliability testing without thoroughly examining item characteristics. By incorporating reliability analysis and item analysis, this study provides stronger empirical evidence supporting the instrument's effectiveness for assessing critical thinking skills in elementary mathematics problem solving.

Practical Is The Assessment Instrument For Critical Thinking Skills In Elementary School Mathematical Problem Solving Based On Evaluations From Experts And Users (Teachers)

The fifth research problem in this study focuses on analyzing the practicality level of the developed instrument for assessing critical thinking skills in elementary school mathematics problem solving based on evaluations from experts and users, particularly teachers. Practicality is an essential component in educational instrument development because an assessment tool, regardless of its validity and reliability, must also be easy to use, understandable, and applicable in real classroom situations. In the context of elementary education, teachers often face challenges related to limited instructional time, administrative workload, and varying levels of assessment literacy. Therefore, the practicality of an assessment instrument becomes highly important to ensure that teachers can implement it effectively without disrupting the teaching and learning process. The thesis emphasizes that the developed instrument was not only intended to meet psychometric standards but also designed to support authentic and formative assessment practices aligned with the principles of the Kurikulum Merdeka.

The results presented in the thesis indicate that the instrument achieved a very high level of practicality. Based on expert evaluations, the practicality score reached 88.13%, while teacher responses showed a practicality level of 85.94%. Furthermore, observations of implementation produced a score of 85.16%, resulting in an overall average practicality level of 86.41%. These findings demonstrate that the instrument is categorized as highly practical for classroom use. In educational research, practicality percentages above 80% are generally considered indicators of excellent usability and feasibility. Therefore, the results suggest that both experts and teachers

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

perceived the instrument as clear, manageable, and suitable for assessing students' critical thinking skills in mathematics learning.

One important factor contributing to the instrument's practicality is its systematic design. The instrument consists of contextual essay questions, detailed scoring rubrics, clear instructions, and assessment guidelines that facilitate implementation by teachers. The use of analytical scoring rubrics is particularly significant because it helps teachers evaluate students' responses objectively and consistently. Essay-based assessments are often criticized for requiring extensive time and involving subjective judgment; however, the inclusion of detailed scoring criteria in this instrument appears to have minimized these challenges. The rubrics provide explicit descriptions of performance levels, ranging from students who demonstrate deep understanding and logical reasoning to those who provide incomplete or irrelevant responses. Consequently, teachers can assess students' critical thinking processes more systematically and accurately.

Another important aspect of practicality highlighted in the thesis is the contextual nature of the assessment tasks. The instrument uses real-life mathematical situations familiar to elementary school students, such as fractions, measurements, and everyday numerical problems. This contextual approach not only supports authentic assessment principles but also makes the instrument easier for teachers to integrate into regular classroom learning activities. Because the problems are closely related to students' daily experiences, teachers reported that students were more engaged and motivated during the assessment process. This finding aligns with educational theories emphasizing that contextualized learning enhances students' understanding and participation in mathematical reasoning activities.

The practicality analysis also reflects the relevance of the instrument to current curriculum demands. The Kurikulum Merdeka encourages teachers to implement formative and diagnostic assessments that focus on students' competencies and thinking processes rather than solely measuring factual knowledge. However, many teachers still struggle to develop assessment instruments capable of evaluating higher-order thinking skills systematically. The developed instrument addresses this challenge by providing a ready-to-use assessment tool grounded in critical thinking theory and mathematical problem-solving principles. Teachers can therefore use the instrument not only to assess students' learning outcomes but also to identify strengths and weaknesses in students' reasoning processes. This diagnostic function is highly valuable because it enables teachers to design more targeted and differentiated instructional strategies.

The study further reveals that teachers perceived the instrument as beneficial for improving the quality of mathematics assessment. Many traditional mathematics tests primarily

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

focus on procedural calculations and correct answers, whereas this instrument emphasizes reasoning, explanation, and evaluation. Teachers reported that the instrument allowed them to gain deeper insights into students' thinking processes, particularly in understanding how students interpret problems, devise strategies, and justify solutions. This shift from product-oriented assessment to process-oriented assessment is consistent with modern educational perspectives emphasizing the development of higher-order thinking skills.

In addition, the practicality findings demonstrate that the ADDIE development model used in the study contributed significantly to the usability of the final product. Through systematic stages of analysis, design, development, implementation, and evaluation, the instrument underwent continuous revisions based on expert suggestions and field-testing results. This iterative process ensured that the final product was not only theoretically sound but also practically applicable in elementary school contexts. The involvement of teachers during the implementation and evaluation stages further strengthened the instrument's relevance to classroom realities.

CONCLUSION

Based on the results and discussion of the study, it can be concluded that the assessment instrument for critical thinking skills in mathematical problem solving at the elementary school level is highly needed to support twenty-first-century competencies. The instrument is designed not only to assess students' final answers but also to reveal their thinking processes, including understanding problems, planning strategies, implementing solutions, and reviewing results. The instrument was systematically developed using the ADDIE model through the stages of analysis, design, development, implementation, and evaluation to ensure the quality of the final product. The findings showed that the developed instrument possessed strong content and construct validity, indicating that the assessment items appropriately represented the dimensions of critical thinking skills theoretically and empirically. In addition, the instrument demonstrated high reliability, moderate difficulty levels, and adequate discrimination indices, confirming its effectiveness in measuring variations in students' abilities accurately. The practicality evaluation from experts and teachers also categorized the instrument as highly practical because it was easy to use, supported by clear instructions, and equipped with analytical scoring rubrics that facilitated objective and consistent scoring. Therefore, the developed instrument fulfills the requirements of validity, reliability, and practicality, making it highly applicable for evaluating critical thinking skills in elementary mathematics learning.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

FUTURE DIRECTIONS

Future research is recommended to expand the implementation of the developed critical thinking assessment instrument in broader educational settings involving larger and more diverse student populations across different regions and grade levels. Further studies may also integrate digital or technology-based assessment systems to improve efficiency, accessibility, and real-time feedback in mathematics learning. In addition, future researchers are encouraged to explore the effectiveness of the instrument in measuring other higher-order thinking skills, such as creative thinking and problem-solving collaboration, as well as examine its impact on students' long-term mathematical achievement and learning motivation. Continuous refinement of assessment indicators and scoring rubrics is also important to ensure alignment with evolving curriculum demands and twenty-first-century educational competencies.

REFERENCES

Arini, N. W., Sudarma, I. K., & Suartama, I. K. (2020). Development of critical thinking assessment instruments for elementary school students. Jurnal Pendidikan Dasar Indonesia, 4(3), 487–495.

Astiwi, N. P. W., Putra, D. K. N. S., & Ganing, N. N. (2020). Development of higher-order thinking skills assessment in elementary mathematics learning. International Journal of Elementary Education, 4(4), 458–466.

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), The Blackwell guide to the philosophy of education (pp. 181–193). Blackwell Publishing.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? CBE—Life Sciences Education, 15(4), 1–7.

Brookfield, S. D. (2013). Teaching for critical thinking: Tools and techniques to help students question their assumptions. Jossey-Bass.

Dewi, N. P., Sariyasa, S., & Putrayasa, I. B. (2020). Development of critical thinking assessment instruments in mathematics learning. Jurnal Pendidikan Matematika, 14(2), 115–128.

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. University of Illinois.

Facione, P. A. (2015). Critical thinking: What it is and why it counts (6th ed.). Insight Assessment.

Feltz, D. L., Chase, M. A., Moritz, S. E., & Sullivan, P. J. (1999). A conceptual model of coaching efficacy: Preliminary investigation and instrument development. Journal of Educational

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Psychology, 91(4), 765–776.
Kalkbrenner, M. T. (2021). The MEASURE approach to instrument development and validation. Measurement and Evaluation in Counseling and Development, 54(3), 177–193.
Kemendikbudristek. (2023). Asesmen Kompetensi Minimum dan karakteristik pembelajaran abad 21. Ministry of Education, Culture, Research, and Technology of Indonesia.
Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. Research in Nursing & Health, 30(2), 224–234.
Marselina, Y., Rahman, A., & Fitriani, N. (2021). Development of mathematical critical thinking assessment instruments for elementary students. Jurnal Evaluasi Pendidikan, 12(1), 45–57.
McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Instrument development in the affective domain: School and corporate applications (3rd ed.). Springer.
Moon, J. A. (2008). Critical thinking: An exploration of theory and practice. Routledge.
Ndiung, S., & Jediut, M. (2020). Development of higher-order thinking skills through mathematical problem-solving learning. Journal of Mathematics Education, 11(3), 221–233.
Organisation for Economic Co-operation and Development. (2019). PISA 2018 results: What students know and can do. OECD Publishing.
Paul, R., & Elder, L. (2014). Critical thinking: Tools for taking charge of your professional and personal life (2nd ed.). Pearson Education.
George Polya. (1973). How to solve it: A new aspect of mathematical method (2nd ed.). Princeton University Press.
Regelski, T. A. (2005). Critical thinking and reflective practice in education. Philosophy of Music Education Review, 13(1), 1–15.
Tavakol, M., & Wetzel, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. International Journal of Medical Education, 11, 245–247.
Yudha, R. (2019). Rasch model analysis in developing critical thinking assessment instruments. Jurnal Penelitian Pendidikan, 19(2), 98–110.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License