JPPIPA 12(11) (2026)

Jurnal Penelitian Pendidikan IPA

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

Five-Tier Diagnostic Test Instrument to Identify Students' Misconceptions and Critical Thinking on Chemical Bonding Material

Putri Suci Anugrah[1*], Roza Linda[1], Rasmiwetti[1]

[1] Postgraduate Chemistry Education Program, Riau University, Pekanbaru, Indonesia.

Received: June 9, 2026
Revised: September 22, 2026
Accepted: November 17, 2026
Published: November 30, 2026

Corresponding Author:
Putri Suci Anugrah
putrisucianugrah@gmail.com

DOI: 10.29303/jppipa.v12i11.15615

Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)

Abstract: Indonesian students still face challenges in developing critical thinking skills and may develop misconceptions in chemistry learning, particularly in chemical bonding. Misconceptions refer to understandings that are inconsistent with scientifically accepted concepts and may hinder subsequent learning. This study aimed to develop a five-tier diagnostic test instrument integrated with critical thinking skills to identify students' misconceptions in chemical bonding material. The study employed a Research and Development (R&D) method using the ADDIE model, consisting of analysis, design, development, implementation, and evaluation stages. The participants were students of Grade XI from SMAN 1 Pekanbaru, SMAN 4 Pekanbaru, and SMAN 14 Pekanbaru, selected using a sampling technique appropriate to the school-based research design. Instrument validation involved three chemistry material experts, followed by limited and large-scale trials. The results showed that the developed instrument demonstrated good validity and reliability, with 36 of 40 items declared valid and high internal consistency. User responses also indicated that the instrument was highly feasible for use. The diagnostic results revealed differences in students' conceptual understanding, misconceptions, and critical thinking skills across school categories, with students in the high-category school demonstrating the strongest conceptual understanding and critical thinking performance. ANOVA results confirmed a significant difference in critical thinking ability among the school categories. These findings indicate that the developed five-tier diagnostic instrument is suitable for identifying misconceptions while simultaneously assessing students' critical thinking skills, thereby providing teachers with more comprehensive diagnostic information for improving chemistry instruction.

Keywords: Chemical bonds; Critical thinking; Five-tier; Misconception

Introduction

The development of quality human resources in education has shifted its orientation to meet 21st-century competencies (Maritsa et al., 2021) . According to Yunus and Mitrohardjoyo, the world is currently moving towards the era of Society 5.0. The arrival of Society 5.0 has made a significant contribution to Indonesian education. This era emphasizes readiness to develop creativity and critical thinking (Teknowijoyo, 2022).

Critical thinking skills are urgent for students, this is because critical thinking skills are able to make students think logically and determine the best choices for themselves (Zakiah & Lestari, 2019). In addition, this skill functions to solve problems (Ma et al., 2021). A critical thinker is able to search, understand, and evaluate statements logically and rationally in the process of problem solving and decision making (Shaw et al., 2020). Critical thinking skills influence reflective thinking and reasoning to choose what to believe (Ennis, 2018). Students who have critical thinking skills do not necessarily agree with the teacher's explanation, because these students can build on the knowledge they have gained by mastering all concepts (Doyan et al., 2020,

How to Cite:
Anugrah, P. S., Linda, R., & Rasmiwetti, R. (2026). Five-Tier Diagnostic Test Instrument to Identify Students' Misconceptions and Critical Thinking on Chemical Bonding Material. *Jurnal Penelitian Pendidikan IPA*, *12*(11), 49–63. https://doi.org/10.29303/jppipa.v12i11.15615

2024; Muliyadi, Doyan, et al., 2026; Rosyana et al., 2019). Based on research by Saputra et al. (2018) stated that critical thinking is a 21st-century skill that ranks first with a percentage of 78% in the list of required skills.

Indonesian students still have low critical thinking skills. This is because students are unable to explain and solve problems comprehensively (Widiandari & Redhana, 2021). Yunita et al. (2018) explain that through test instruments, students' critical thinking skills can be developed. Students' critical thinking skills can be measured using test instruments prioritized to measure these skills by focusing on the aspects contained therein (Laksono, 2020). Critical thinking skills can also be measured using diagnostic test instruments (Doyan et al., 2026; Kartini et al., 2019; Muliyadi, Hamidi, et al., 2026). Diagnostic tests are test designs that function to identify student weaknesses. Thus, the results can be used as a basis for determining appropriate follow-up treatment for students in the future (Prianti et al., 2020).

The results of the PISA (Planning for International Student Assessment) conducted by the Organization for Economic Co-operation and Development (OECD) show that Indonesian students' critical thinking skills remain low, as evidenced by Indonesia's ranking in the bottom group from 2009 to 2018 (OECD, 2019). This condition indicates that students still struggle to identify, understand, and apply scientific knowledge to solve more complex problems (Cahyana et al., 2019; Hawa & Putra, 2018).

Therefore, teachers need to conduct learning evaluations to determine students' mastery of the material and learning difficulties (Hidayah & Muchtar, 2022). In chemistry learning, conceptual understanding is crucial because chemical concepts are interrelated and hierarchical. Misunderstanding basic concepts can lead to difficulties in understanding more complex concepts and lead to misconceptions (Hidayah & Muchtar, 2022).

Chemistry is a branch of science that studies matter, its structure, properties, and the changes and reactions that accompany them. Concepts in chemistry are interrelated and hierarchical, so that errors in understanding basic concepts can cause difficulties in understanding more complex concepts (Rokhim, 2023). One of the goals of chemistry learning is for students to be able to understand and apply concepts appropriately in accordance with Permendikbud Number 21 of 2016 (Zainab & Sukarmin, 2020). Misunderstanding concepts can cause learning difficulties, low learning outcomes, and hinder the acceptance of new knowledge (Hidayati, 2019). In chemistry learning, it is often found that students have concepts that differ from those accepted by experts or scientifically. This mismatch in conceptual understanding is also called misconception (Suparno, 2013). In chemistry, misconceptions are a serious problem because the concepts are hierarchical and interrelated, so that misconceptions in one concept can affect the understanding of the next concept (Savira, 2019).

Based on a literature review conducted by Rokhim (2023), identified chemical bonding as one of the chemistry topics that frequently involves misconceptions. Interviews with chemistry teachers at SMAN 1, SMAN 4, and SMAN 14 Pekanbaru also indicated that students experience difficulties in understanding chemical bonding, particularly electronegativity, electron configuration, Lewis structures, and types of chemical bonds. The abstract and interconnected nature of these concepts can lead to misconceptions and difficulties in subsequent learning (Suwarto, 2013). Therefore, an appropriate diagnostic assessment instrument is needed to identify students' misconceptions in chemical bonding accurately.

Diagnostic assessment instruments are used to Diagnostic assessment instruments play an important role in identifying students' conceptual understanding, learning difficulties, and misconceptions. Conventional multiple-choice assessments, however, are generally limited in distinguishing between students who understand a concept, students who guess the correct answer, and students who hold an incorrect conception with high confidence. Multi-tier diagnostic tests address this limitation by combining students' answers with additional information about their confidence and reasoning. One of these formats is the five-tier diagnostic test, which consists of the answer to the question, confidence in the answer, the reason for the answer, confidence in the reason, and the source of information used by the student (Putra, 2020). The combination of these components enables a more comprehensive diagnosis of students' conceptual status and provides information that can help teachers identify misconceptions and their possible sources.

The five-tier multiple choice diagnostic test has the advantage of being able to identify students' misconceptions in more depth, find out the causes of misconceptions, determine material that needs to be emphasized, and help plan more effective learning.

Previous studies have shown that multi-tier diagnostic tests are effective in identifying misconceptions. Rosita (2020) developed a Five-Tier Newton's Laws Test, while (Rumapea & Silaban, 2022) developed a Three-Tier diagnostic test. However, these studies primarily focused on misconceptions and did not explicitly integrate critical thinking skills into the diagnostic instrument. Thus, previous research has not sufficiently addressed the need for an instrument that simultaneously identifies misconceptions and assesses critical thinking, particularly in chemical bonding material.

Preliminary interviews with chemistry teachers at SMAN 1 Pekanbaru, SMAN 4 Pekanbaru, and SMAN 14 Pekanbaru indicated that diagnostic assessment was still limited to oral questioning and interviews, while written assessments were predominantly focused on lower-order cognitive levels (C1–C3). Consequently, students' misconceptions and critical-thinking abilities could not be identified comprehensively.

The novelty of this study lies in developing a Five-Tier Diagnostic Test Instrument for chemical bonding that integrates misconception identification with critical-thinking assessment. The instrument is expected to distinguish students' conceptual understanding, misconceptions, and reasoning abilities through their answers, confidence, reasons, and information sources.

Based on these problems and research gaps, this study aims to develop and determine the quality of a Five-Tier Diagnostic Test Instrument to Identify Students' Misconceptions and Critical Thinking on Chemical Bonding Material. The instrument is expected to provide teachers with comprehensive diagnostic information for identifying conceptual and reasoning difficulties and supporting more targeted chemistry instruction.

Method

Research Approach

This study employed a Research and Development (R&D) approach using the ADDIE model, which consists of five stages: Analysis, Design, Development, Implementation, and Evaluation. The purpose of this research was to develop a valid and reliable five-tier diagnostic test instrument integrated with critical thinking skills to identify students' misconceptions on chemical bonding concepts.

Analysis Stage

The analysis phase was conducted to identify research needs. Activities included analyzing the problem, students, and materials through teacher interviews and literature reviews. The results of the analysis were used as the basis for instrument development.

Design Stage

To be developed is designed. Activities include developing a grid, validation instruments, response questionnaires, and test items and answer keys. The design is tailored to indicators of misconceptions and critical thinking skills.

Development Stage

The designed instrument was then developed into a preliminary product. This product was validated by three subject matter experts to assess its feasibility. The validation results were used as the basis for instrument revisions.

Implementation Stage

The revised instrument was tested through individual and small group trials, and field trials. These trials were conducted to determine the instrument's quality and effectiveness. The results of the trials served as the basis for product improvements.

Evaluation Stage

The evaluation phase was conducted after the entire trial process was completed. The data obtained was analyzed to determine the level of misconceptions and critical thinking skills of students. The results of the analysis were then used to draw research conclusions.

Place and Time of Research

The research was conducted at the Postgraduate Program in Chemistry Education, Faculty of Teacher Training and Education, University of Riau. Product trials were conducted at SMAN 1 Pekanbaru, SMAN 4 Pekanbaru, and SMAN 14 Pekanbaru, representing high-, medium-, and low-category schools based on their respective school characteristics and academic conditions. Data collection was conducted from September to October 2024.

Data Types and Sources

The research used quantitative and qualitative data. Quantitative data were obtained from instrument validation, test results, questionnaires, and instrument characterization. Qualitative data were obtained from interviews, expert suggestions, criticisms, and feedback during the instrument development process.

Population and Sample

The population of this study consisted of Grade XI MIA students from public senior high schools in Pekanbaru City. The sample consisted of 10 students from SMAN 1 Pekanbaru, 10 students from SMAN 4 Pekanbaru, and 10 students from SMAN 14 Pekanbaru, with a total of [total sample] students. Purposive sampling was used to select schools based on predetermined characteristics, including school category and suitability with the research objectives and implementation requirements.

Data Collection Technique

Data were collected through interviews, tests, questionnaires, and expert validation. Interviews were conducted to obtain information about students' misconceptions and learning difficulties. Tests and questionnaires were used to assess the quality of the

developed instrument and students' critical thinking skills. The instrument development was conducted through a limited trial, followed by revision based on the results, and then a field trial to evaluate the instrument on a larger scale.

Research Instruments

The primary research instrument was a five-tier diagnostic test consisting of 40 items. Each item consisted of five tiers: (1) the answer to the question, (2) confidence in the answer, (3) the reason for the answer, (4) confidence in the reason, and (5) the source of information used by the student. The instrument was designed to identify students' misconceptions and critical thinking skills on chemical bonding material. The critical thinking indicators included interpretation, analysis, evaluation, inference, explanation, and self-regulation, while misconception identification was based on the consistency of students' answers, reasons, confidence levels, and sources of information.

Supporting instruments included expert validation sheets and teacher and student response questionnaires. The instrument was validated by three chemistry education experts who assessed the content, construction, language, suitability of the items with the chemical bonding material, misconception indicators, and critical thinking indicators. Content validity was analyzed using Aiken's V coefficient. An item was considered valid when the Aiken's V value met the predetermined acceptance criterion of ≥ 0.80. Based on the validation and item analysis, 36 of the 40 developed items were declared valid and suitable for further testing, while the remaining items were revised or excluded. The teacher and student response questionnaires were used to determine the practicality and feasibility of the developed instrument.

Data Analysis Techniques

Data were analyzed using Microsoft Excel and IBM SPSS Statistics. Instrument quality was assessed through validity, reliability, difficulty level, and item discrimination. Misconceptions were identified based on students' answers, confidence levels, reasons, and information sources, while user responses and critical thinking skills were analyzed descriptively based on predetermined criteria.

Differences in critical thinking skills among school categories were analyzed using one-way ANOVA in SPSS. Normality and homogeneity tests were conducted as prerequisites, with a significance level of 0.05. A significance value < 0.05 indicated a significant difference between school categories.

Research permission was obtained from the relevant authorities and participating schools. Respondents' data were kept confidential and used solely for research purposes.

Result and Discussion

Analysis Phase

Problem Analysis

Interviews with chemistry teachers at SMAN 1, SMAN 4, and SMAN 14 Pekanbaru revealed that the questions used were still dominated by lower-level thinking questions and did not measure critical thinking skills. Furthermore, questions were often adopted from various sources without in-depth analysis of learning outcomes. Therefore, it is necessary to develop an instrument capable of measuring students' critical thinking skills.

Needs Analysis

A needs analysis shows that students require assessment instruments that align with the demands of the Independent Curriculum and are capable of measuring critical thinking skills. Based on Jean Piaget's cognitive development theory, high school students are at a stage where they can think abstractly and critically. Therefore, valid, reliable instruments capable of measuring higher-order thinking skills are required.

Material Domain Analysis

Material analysis was conducted on the topic of chemical bonds, referring to the learning outcomes of the Independent Curriculum and various literature related to misconceptions and critical thinking. The instrument was developed in the form of five-tier questions at the C3–C5 cognitive levels to identify misconception categories according to Kaniawati (2019) and measure critical thinking indicators according to Ennis (2011). Thus, the instrument is expected to be able to identify misconceptions and assess students' level of critical thinking on chemical bonds.

Design Phase

Developing a Quality Assessment Instrument for Diagnostic Test Instruments

At this stage, an instrument was developed to assess the quality of the five-tier diagnostic test, a questionnaire aimed at subject matter experts, chemistry teachers, and students. The questionnaire was compiled in checklist form based on a pre-designed grid and then consulted with and validated by lecturers at the University of Riau's Chemistry Education Department. This stage aimed to ensure the developed instrument was suitable for use.

Determination of the Draft Assessment Instrument

This stage resulted in the initial design of an instrument for assessing misconceptions and critical thinking skills in the form of multiple-choice questions. The question format was selected based on its ability to measure higher-order thinking skills more objectively and effectively. The draft instrument was designed according to the learning objectives and the characteristics of the chemical bonding material.

Compiling Questions and Answers

At this stage, questions and answer keys are developed based on learning indicators for chemical bonding. Questions are developed by referring to relevant literature to align with scientific concepts and measurement objectives. The result of this stage is a set of five-tier diagnostic questions ready for validation and testing.

Development Phase

In the development stage, the five-tier misconception and critical thinking ability assessment instrument on chemical bonding material was prepared and then validated by three expert validators. Validation was carried out to assess the suitability of the material, language, construction, and critical thinking ability aspects using the Aiken's V index. The revised instrument based on the validator's suggestions was then tested on students to determine the characteristics of the questions.

Table 1. Validation Criteria Aspects

No	Validation Criteria Aspects
Material Aspect	
1	The relationship between learning outcomes and the flow of learning objectives
2	Alignment of learning objectives with FTMCT diagnostic instruments
3	FTMCT diagnostic materials and instruments
4	The content of the material asked is according to the level, type of school or school level.
Language Aspects	
1	Understanding of messages or information
2	Language presentation in FTMCT diagnostic instruments
3	The suitability of the language used to the level of thinking of students
Construction Aspects	
1	Presentation of images, tables, graphs or similar on diagnostic instruments
2	Readability of FTMCT diagnostic instruments
3	The suitability of the breadth of questions to the development of high school students
4	The suitability of the diagnostic instrument with the answers from the instrument
5	Completeness of assessment instruments (question outline, question items, and answers)

Validation results showed that all aspects met the established criteria. The material aspect ensured the questions were appropriate to the learning objectives, the language aspect assessed the clarity of language use, and the construction aspect assessed the readability and presentation of the questions.

Table 2. Content Validation Results Using Aiken's Formula

No Question	Aiken Validity Index 1	Conclusion	Aiken Validity Index 2	Conclusion
1.	0.729	Quite valid	0.792	Quite valid
2.	0.688	Quite valid	0.854	Very valid
3.	0.660	Quite valid	0.854	Very valid
4.	0.715	Quite valid	0.861	Very valid
5.	0.708	Quite valid	0.868	Very valid
6.	0.681	Quite valid	0.875	Very valid
7.	0.708	Quite valid	0.875	Very valid
8.	0.771	Quite valid	0.882	Very valid
9.	0.729	Quite valid	0.882	Very valid
10.	0.750	Quite valid	0.889	Very valid
11.	0.715	Quite valid	0.889	Very valid
12.	0.785	Quite valid	0.889	Very valid
13.	0.708	Quite valid	0.896	Very valid
14.	0.722	Quite valid	0.903	Very valid
15.	0.722	Quite valid	0.910	Very valid
16.	0.743	Quite valid	0.910	Very valid
17.	0.715	Quite valid	0.910	Very valid
18.	0.778	Quite valid	0.910	Very valid

No Question	Aiken Validity Index 1	Conclusion	Aiken Validity Index 2	Conclusion
19.	0.736	Quite valid	0.910	Very valid
20.	0.750	Quite valid	0.910	Very valid
21.	0.736	Quite valid	0.917	Very valid
22.	0.778	Quite valid	0.917	Very valid
23.	0.708	Quite valid	0.917	Very valid
24.	0.750	Quite valid	0.917	Very valid
25.	0.736	Quite valid	0.917	Very valid
26.	0.729	Quite valid	0.917	Very valid
27.	0.688	Quite valid	0.924	Very valid
28.	0.660	Quite valid	0.924	Very valid
29.	0.715	Quite valid	0.924	Very valid
30.	0.708	Quite valid	0.924	Very valid
31.	0.681	Quite valid	0.938	Very valid
32.	0.708	Quite valid	0.938	Very valid
33.	0.771	Quite valid	0.938	Very valid
34.	0.729	Quite valid	0.938	Very valid
35.	0.750	Quite valid	0.944	Very valid
36.	0.715	Quite valid	0.944	Very valid
37.	0.785	Quite valid	0.944	Very valid
38.	0.708	Quite valid	0.944	Very valid
39.	0.722	Quite valid	0.965	Very valid
40.	0.722	Quite valid	0.965	Very valid

Table 3. Percentage Results of Critical Thinking Measurement Validation

Validator	Total Validation Score 1	Total Validation Score 2
I	29%	40%
II	31%	36%
III	28%	37%
Score	88%	113%
%	73.3%	94.2%

Based on the content validation results, the Aiken's V value in the first stage was in the range of 0.66–0.78, so some questions still needed revision. After improvements were made according to the validator's suggestions, the Aiken's V value increased to 0.79–0.96, indicating that all questions met the validity criteria and were suitable for use.

Validation of the critical thinking ability aspect increased from 73.3 % to 94.2%, categorized as very high. These results indicate that the instrument is suitable for measuring students' critical thinking abilities.

Table 4. Suggestions and Input for Improvement of Each Validator

Validator	Suggestions/Input	Follow-up
Validator 1	Writing reactions to questions should be packaged neatly	Tidying up the exam paper
	Objective arrangement in the form of numbers in order	Improvements were made to the order of the questions.
	The presentation of language in the instrument is not in accordance with EYD, for example spelling errors	Improvements were made to the presentation of language in the instrument according to EYD, for example spelling errors.
Validator 2	Match the indicators with the question indicators on the question grid.	Correction of question indicators to the question grid
	Pay attention to the cognitive level of question no. 2, in question no. 2 there is improvement because it is still at a low level and not yet critical thinking.	Improvements were made to question number 2 to make it a more critical thinking question.
	The question sentence must be adjusted to the writing rules according to EYD (enhanced spelling).	The question sentence is adjusted to the EYD writing rules
Validator 3	Group and arrange questions based on learning objectives	Improvements were made by grouping questions based on GPA.
	critical thinking indicators for each question	Improvements were made to the critical thinking indicators to suit the questions.
	Check again and understand the questions and the appropriateness of the answers	Correct and understand questions regarding the suitability of the answers

In general, the validators provided input regarding grammatical improvements, the appropriateness of indicators to questions, grouping questions based on learning objectives, and enhancing critical thinking

elements in some questions. All of these suggestions were used as the basis for revisions before the instrument was declared valid.

Initial Trial

The validated instrument was piloted on 30 students from three high schools in Pekanbaru. This pilot test aimed to determine construct validity, reliability, difficulty level, and item discrimination.

Construct Validity

The analysis results show that most of the questions have a calculated r value greater than the table r (0.361) and are therefore declared valid. There are several questions that do not meet the validity criteria and need to be eliminated or revised.

Table 5. Validity Test Results Using SPSS

No Question	R Count	R Table	Conclusion
1.	0.464	0.361	Valid
2.	0.412	0.361	Valid
3.	0.427	0.361	Valid
4.	0.561	0.361	Valid
5.	0.490	0.361	Valid
6.	0.629	0.361	Valid
7.	0.462	0.361	Valid
8.	0.411	0.361	Valid
9.	0.362	0.361	Valid
10.	0.655	0.361	Valid
11.	0.534	0.361	Valid
12.	0.628	0.361	Valid
13.	0.497	0.361	Valid
14.	0.597	0.361	Valid
15.	0.434	0.361	Valid
16.	0.434	0.361	Valid
17.	0.417	0.361	Valid
18.	0.340	0.361	Invalid
19.	0.615	0.361	Valid
20.	0.452	0.361	Valid
21.	0.449	0.361	Valid
22.	0.525	0.361	Valid
23.	0.458	0.361	Valid
24.	0.468	0.361	Valid
25.	0.402	0.361	Valid
26.	0.464	0.361	Valid
27.	0.297	0.361	Invalid
28.	0.039	0.361	Invalid
29.	0.561	0.361	Valid
30.	0.490	0.361	Valid
31.	0.629	0.361	Valid
32.	0.462	0.361	Valid
33.	0.216	0.361	Invalid
34.	0.428	0.361	Valid
35.	0.655	0.361	Valid
36.	0.534	0.361	Valid
37.	0.628	0.361	Valid
38.	0.497	0.361	Valid
39.	0.597	0.361	Valid
40.	0.434	0.361	Valid

Reliability Test

The reliability test results yielded a Cronbach's Alpha value of 0.875. This value indicates that the instrument has a very good level of consistency, making it suitable for use as a measuring tool.

Table 6. Reliability Test Results

No Question	Cronbach's Alpha if Item Deleted	Cronbach's Alpha	Information
1	0.911	0.875	Reliable
2	0.911	0.875	Reliable
3	0.911	0.875	Reliable
4	0.91	0.875	Reliable
5	0.911	0.875	Reliable
6	0.909	0.875	Reliable
7	0.911	0.875	Reliable
8	0.912	0.875	Reliable
9	0.912	0.875	Reliable
10	0.908	0.875	Reliable
11	0.91	0.875	Reliable
12	0.909	0.875	Reliable
13	0.911	0.875	Reliable
14	0.909	0.875	Reliable
15	0.911	0.875	Reliable
16	0.911	0.875	Reliable
17	0.912	0.875	Reliable
18	0.912	0.875	Reliable
19	0.909	0.875	Reliable
20	0.911	0.875	Reliable
21	0.911	0.875	Reliable
22	0.91	0.875	Reliable
23	0.911	0.875	Reliable
24	0.911	0.875	Reliable
25	0.912	0.875	Reliable
26	0.911	0.875	Reliable
27	0.913	0.875	Reliable
28	0.916	0.875	Reliable
29	0.91	0.875	Reliable
30	0.911	0.875	Reliable
31	0.909	0.875	Reliable
32	0.911	0.875	Reliable
33	0.914	0.875	Reliable
34	0.911	0.875	Reliable
35	0.908	0.875	Reliable
36	0.91	0.875	Reliable
37	0.909	0.875	Reliable
38	0.911	0.875	Reliable
39	0.909	0.875	Reliable
40	0.911	0.875	Reliable

Level of Difficulty of Questions

Analysis of the difficulty level shows that the questions are divided into easy, medium, and difficult categories. Most of the questions fall into the medium category, making them suitable for optimally measuring student ability. The question set consists of 6 easy questions, 29 medium questions, and 5 difficult questions.

Table 7. Results of Calculation of Question Difficulty Level

No Question	Mark	Category
1	0.6	currently
2	0.83	easy
3	0.57	currently
4	0.73	easy
5	0.7	currently
6	0.6	currently
7	0.73	easy
8	0.67	currently
9	0.23	difficult
10	0.53	currently
11	0.5	currently
12	0.47	currently
13	0.5	currently
14	0.33	currently
15	0.57	currently
16	0.57	currently
17	0.5	currently
18	0.3	difficult
19	0.5	currently
20	0.83	easy
21	0.6	currently
22	0.47	currently
23	0.3	difficult
24	0.5	currently
25	0.5	currently
26	0.6	currently
27	0.43	currently
28	0.3	difficult
29	0.73	easy
30	0.7	currently
31	0.6	currently
32	0.73	easy
33	0.3	difficult
34	0.53	currently
35	0.53	currently
36	0.5	currently
37	0.47	currently
38	0.5	currently
39	0.33	currently
40	0.57	currently

Distinguishing Power

The results of the discriminatory power analysis showed that the majority of questions were categorized as good to excellent. Twenty-four questions were categorized as excellent and six were categorized as good, indicating that the instrument was able to distinguish between high-ability and low-ability students.

Table 8. Valid Distinguishing Power Values of Question Items

No Question	Differentiator	Category
1	0.27	Enough
2	0.33	Good
3	0.33	Good
4	0.53	Very good
5	0.47	Very good
6	0.53	Very good
7	0.27	Enough
8	0.27	Enough
9	0.33	Good
10	0.67	Very good
11	0.47	Very good
12	0.53	Very good
13	0.33	Good
14	0.4	Very good
15	0.6	Very good
16	0.47	Very good
17	0.33	Good
18	0.47	Very good
19	0.47	Very good
20	0.2	Enough
21	0.4	Very good
22	0.53	Very good
23	0.4	Very good
24	0.47	Very good
25	0.2	Enough
26	0.27	Enough
27	0.2	Enough
28	0	not good
29	0.53	Very good
30	0.47	Very good
31	0.53	Very good
32	0.27	Enough
33	0.07	not good
34	0.4	Very good
35	0.67	Very good
36	0.47	Very good
37	0.53	Very good
38	0.33	Good
39	0.4	Very good
40	0.6	Very good

Table 9. Characteristics of the Five-tier Misconception Assessment Instrument

No Question	Validity	Reliability	Difficulty of Questions	Differential Power	Category
1	0.464	0.911	0.60	0.27	Qualify
2	0.412	0.911	0.83	0.33	Qualify
3	0.427	0.911	0.57	0.33	Qualify
4	0.561	0.910	0.73	0.53	Qualify
5	0.490	0.911	0.70	0.47	Qualify
6	0.629	0.909	0.60	0.53	Qualify
7	0.462	0.911	0.73	0.27	Qualify
8	0.411	0.912	0.67	0.27	Qualify

No Question	Validity	Reliability	Difficulty of Questions	Differential Power	Category
9	0.362	0.912	0.23	0.33	Qualify
10	0.655	0.908	0.53	0.67	Qualify
11	0.534	0.910	0.50	0.47	Qualify
12	0.628	0.909	0.47	0.53	Qualify
13	0.497	0.911	0.50	0.33	Qualify
14	0.597	0.909	0.33	0.40	Qualify
15	0.434	0.911	0.57	0.60	Qualify
16	0.434	0.911	0.57	0.47	Qualify
17	0.417	0.912	0.50	0.33	Qualify
18	0.340	0.912	0.30	0.47	Not eligible
19	0.615	0.909	0.50	0.47	Qualify
20	0.452	0.911	0.83	0.20	Qualify
21	0.449	0.911	0.60	0.40	Qualify
22	0.525	0.910	0.47	0.53	Qualify
23	0.458	0.911	0.30	0.40	Qualify
24	0.468	0.911	0.50	0.47	Qualify
25	0.402	0.912	0.50	0.20	Qualify
26	0.464	0.911	0.60	0.27	Qualify
27	0.297	0.913	0.43	0.20	Not eligible
28	0.039	0.916	0.30	0.00	Not eligible
29	0.561	0.910	0.73	0.53	Qualify
30	0.490	0.911	0.70	0.47	Qualify
31	0.629	0.909	0.60	0.53	Qualify
32	0.462	0.911	0.73	0.27	Qualify
33	0.216	0.914	0.30	0.07	Not eligible
34	0.428	0.911	0.53	0.40	Qualify
35	0.655	0.908	0.53	0.67	Qualify
36	0.534	0.910	0.50	0.47	Qualify
37	0.628	0.909	0.47	0.53	Qualify
38	0.497	0.911	0.50	0.33	Qualify
39	0.597	0.909	0.33	0.40	Qualify
40	0.434	0.911	0.57	0.60	Qualify

Based on the overall analysis of the instrument's characteristics, 36 items were obtained that met the requirements for validity, reliability, difficulty level, and discriminatory power. These items were then used in the next testing stage.

One-on-One Trial

One-on-one trials were conducted with three students with high, medium, and low abilities. This activity aimed to determine readability, clarity of instructions, and time spent completing the questions.

Table 10. One-on-One Trial of the Five-tier Misconception Instrument

Category	Processing time (minutes)	Score
Tall	80	100
Currently	90	95
Low	110	80

Test results showed that high-ability students completed the questions faster than other students. Overall, students reported that the questions were easy to understand, visually appealing, and demonstrated critical thinking skills. Therefore, the instrument was deemed suitable for use with minor revisions.

Limited Trial

A limited trial was conducted on 10 students and 3 chemistry teachers to determine the readability, time adequacy, and suitability of the instrument.

Student Response

The questionnaire results showed that the readability aspect was categorized as very good, while the time adequacy aspect was categorized as good. Students rated the language used as easy to understand, the question structure clear, and the time given to complete the task sufficient.

Table 11. Results of Student Response Analysis

Aspect	Mark	Category
Readability of Questions	91	Very good
Time Sufficiency	71	Good

Teacher's Response

The teacher questionnaire results showed an instrument suitability rating of 84%, categorized as very

good. Teachers assessed that the material aligned with the learning objectives, the language used was communicative, and the questions encouraged students to think critically and solve problems. Based on these results, the instrument was acceptable and could be used as the final product without further revision.

Implementation Phase

The implementation phase is the stage of product use on a large scale. This activity includes learning observations and assessments using a five-tier misconception instrument on chemical bonding. The instrument has undergone quality testing, including validity, reliability, difficulty level, and discriminatory power. The instrument serves as an assessment of learning to measure student learning outcomes after the learning process.

The large-scale test was conducted on 90 grade XI IPA students from SMAN 1 Pekanbaru, SMAN 4 Pekanbaru, and SMAN 14 Pekanbaru, each with 30 students in the even semester of the 2024/2025 academic year.

Misconception

The results of identifying students' misconceptions based on five levels of conception, namely misconception (MC), understanding the concept (SU), partially understanding the concept (PU), not understanding (NU), and not defined (UC) are presented in Table 12.

Table 12. Student Identification Results

School	MC	SU	Public Works	NU	UC
Tall	11.38	60.00	28.00	0.62	0.00
Currently	26.07	43.45	29.93	0.55	0.00
Low	20.91	44.55	34.18	0.36	0.00
Average	19.45	49.33	30.70	0.51	0.00

Based on Table 12, high category schools have the highest percentage of conceptual understanding (SU) at 60.00 % and the lowest misconception (MC) at 11.38%. Conversely, medium category schools have the highest level of misconception at 26.07%. Low category schools show the highest percentage of partial conceptual understanding (PU) at 34.18 %. The percentage of not understanding (NU) is very low in all school categories and no students were found in the undefined category (UC). These findings indicate that students in high category schools have better conceptual understanding than those in medium and low category schools.

Table 13. Percentages Based on Five Levels of Conception

Category	Indicator	Percentage (%)		
		High School	Middle School	Elementary School
Misconception (MC)	MC-B	1.23	1.79	1.27
	MC-T	2.26	0.41	1.27
	MC-PT	5.74	9.38	8.18
	MC-OPE	1.64	7.31	4.55
	MC-I	0.51	7.17	5.64
Understand the concept	SUB	5.23	1.93	2.91
(SU)	SU-T	16.00	0.28	5.82
	SU-PT	24.62	13.10	18.00
	SU-OPE	10.87	11.31	7.82
	SU-I	3.28	16.83	10.00
Understand some	PU-B	3.69	2.48	3.27
concepts (PU)	PU-T	4.62	0.14	2.18
	PU-PT	8.51	9.38	11.64
	PU-OPE	8.82	8.83	7.09
	PU-I	2.36	9.10	10.00
Don't understand (NU)	NU-B	0.10	0.00	0.18
	NU-T	0.21	0.00	0.00
	NU-PT	0.10	0.55	0.00
	NU-OPE	0.10	0.00	0.18
	NU-I	0.10	0.00	0.00
Undefined	UC	0.00	0.00	0.00

Based on Table 13, the largest sources of misconceptions come from personal thoughts (MC-PT), other people's explanations (MC-OPE), and the internet (MC-I), especially in medium-category schools. In the conceptual understanding (SU) category, high schools showed the highest percentage, especially in the SU-PT indicator at 24.62 % . Meanwhile, low-category schools had a relatively higher percentage of partial conceptual understanding (PU) compared to other school categories. In general, high-category schools showed

better conceptual understanding, while medium-category schools had higher levels of misconceptions. These results demonstrate the importance of identifying sources of misconceptions so that teachers can design more effective learning strategies to improve students' conceptual understanding.

Critical Thinking Skills

Critical thinking skills were measured using an instrument developed and analyzed by school category. The measurement results are presented in Table 14.

Table 14. Percentage of Critical Thinking Skills in Each School Category

School	Percentage (%)	Category
Tall	80.0	Very high
Currently	68.0	Tall
Low	66.4	Tall

Based on Table 14, students in high-skill schools have the highest critical thinking skills, with a percentage of 80.0%, and are in the very high category. Meanwhile, medium- and low-skill schools are in the high category, with percentages of 68.0% and 66.4%, respectively.

Table 15. Average Critical Thinking Skills

Critical Thinking Ability Indicators	Mark %	Category
Analyzing arguments Analyze cause-and-effect relationships Clarifying information	87	Very high
Inducing and considering the results of induction Identifying relevant information Analyze cause-and-effect relationships	89	Very high
Analyzing arguments Interpreting information from data Determining the implications of the conclusion	67	tall
Analyzing arguments Determining errors or fallacies in thinking Identifying the assumptions underlying the argument	91	Very high
Inducing and considering the results of induction Clarifying and organizing information Analyze cause-and-effect relationships	83	Very high
Analyzing arguments Determining the consistency of arguments Identifying relevant information	70	tall
Analyze cause-and-effect relationships Inducing and considering the results of induction Interpreting information from data	59	currently
Analyze cause-and-effect relationships Identifying the assumptions underlying the argument Identifying relevant information	68	tall
Analyzing arguments Interpreting information from data Clarifying information	76	tall
Determining the results of induction Analyzing arguments Clarifying information	62	tall
Identifying cause-effect patterns Determining the results of induction Analyze cause-and-effect relationships	70	tall
Identifying patterns in data Summarizing the results of the analysis Identifying relevant information	57	currently
Analyzing arguments Applying concepts to answer questions Analyze cause-and-effect relationships	51	currently
Inducing and considering the results of induction Analyzing arguments Analyzing arguments	57	currently
Determining errors or fallacies in thinking		

Critical Thinking Ability Indicators	Mark %	Category
Identifying the assumptions underlying the argument Analyzing arguments Identifying relevant information	67	tall
Determining the implications of the conclusion Analyzing arguments Identifying relevant information	84	Very high
Summarizing the results of the analysis Analyze cause-and-effect relationships Identifying the assumptions underlying the argument	83	Very high
Determining errors or fallacies in thinking Analyzing arguments Interpreting information from data	76	tall
Clarifying information Analyze cause-and-effect relationships Inducing and considering the results of induction	70	tall
Clarifying information Analyzing arguments Identifying patterns in data	79	tall
Summarizing the results of the analysis Analyzing arguments Determining the implications of the conclusion	71	tall
Inducing and considering the results of induction Analyzing arguments Identifying patterns in data	66	tall
Interpreting information from data Analyzing arguments Identifying the assumptions underlying the argument	58	currently
Determining the implications of the conclusion Analyze cause-and-effect relationships Inducing and considering the results of induction	59	currently
Clarifying information Analyzing arguments Identifying patterns in data	84	Very high
Interpreting information from data		

Based on Table 15, most critical thinking skill indicators are in the high to very high category. The indicators with the highest scores are the ability to analyze arguments, identify fallacies, and analyze cause-and-effect relationships, with percentages reaching 91%. Conversely, indicators related to induction and drawing conclusions remain in the moderate category, with percentages ranging from 51% to 59%.

In general, these results indicate that students have developed good critical thinking skills, particularly in analyzing information and establishing cause-and-effect relationships. However, their ability to induce, interpret data, and draw conclusions still needs to be improved through instruction that emphasizes higher-order thinking skills.

Normality Test Results

Based on Table 16, all data have a significance value > 0.05 so that the critical thinking ability data in the three school categories are normally distributed.

Table 16. Large-Scale Data Normality Test

Category	Statistics	Df	Sig.
High School	0.157	30	0.058
Middle School	0.143	30	0.119
Elementary School	0.143	30	0.120

Homogeneity Test Results

The results of the homogeneity test showed a significance value of 0.769 (>0.05), so the data came from a homogeneous group and met the requirements for the ANOVA test.

Table 17. Large-Scale Data Homogeneity Test

Levene Statistics	df1	df2	Sig.
0.263	2	87	0.769

Results of the Critical Thinking Ability Difference Test

Based on the ANOVA results in Table 18, a significance value of 0.000 (< 0.05) was obtained. Thus, there is a significant difference in critical thinking skills between students in high, medium, and low-category schools.

Table 18. One-Way ANOVA Test Results for Large-Scale Data

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	3113,600	2	1556,800	11,891	0.000
Within Groups	11390,400	87	130,924		
Total	14504,000	89			

Table 19. Analysis of Differences in Critical Thinking Skills

School	Comparison	Significance
Tall	Currently	0.001
Tall	Low	0.000
Currently	Tall	0.001
Currently	Low	0.851
Low	Tall	0,000
Low	Currently	0.851

Further test results showed a significant difference between high-skill schools and medium- and low-skill schools ($p < 0.05$). However, there was no significant difference between medium- and low-skill schools ($p = 0.851 > 0.05$). These findings indicate that students' critical thinking skills in high-skill schools were better than those in the other two categories.

Evaluation Phase

The evaluation phase was conducted to assess the suitability of the developed process and product. The results showed that the five-tier misconception assessment instrument, integrated with critical thinking skills on chemical bonding, met the criteria for validity, reliability, and appropriateness for use. The instrument was developed based on the learning outcomes of the Independent Curriculum, with indicators of five-tier misconceptions and critical thinking skills at the cognitive levels C4 (analyzing) and C5 (evaluating).

Of the 40 items developed, 36 were declared valid with a correlation value > 0.36. All items had a Cronbach Alpha value > 0.60, thus being reliable. The results of the analysis of the level of difficulty and discriminating power showed that all items met the established criteria. In addition, the results of the one-on-one and limited trials showed aspects of very good readability (91%), good adequacy (71%), and very good feasibility (84%). Thus, the 36 items were suitable for use in large-scale tests.

High-stakes schools had a higher percentage of conceptual understanding and lower misconceptions than medium- and low-stakes schools. Misconceptions stemming from personal thoughts, other people's explanations, and the internet were more common in medium-stakes schools, while low-stakes schools showed a higher percentage of partial conceptual understanding. These findings suggest that information sources and the way students construct understanding significantly influence the emergence of misconceptions.

The results of critical thinking ability measurements show that students in high-skill schools achieved an average critical thinking ability of 80% (very high), while those in medium- and low-skill schools achieved 68% and 66.4% (high), respectively. The indicators with the highest achievement were the ability to analyze arguments, clarify information, and identify errors in thinking, with percentages ranging from 83% to 91%. Meanwhile, indicators related to the induction process and drawing conclusions still showed lower achievement, requiring more intensive training.

The results of the One Way ANOVA test showed a significant difference in critical thinking skills between the three school categories (sig. = 0.000 < 0.05). Further analysis showed that high category schools were significantly different from medium and low category schools, while medium and low category schools did not show a significant difference. This difference can be influenced by various factors, such as students' academic abilities, psychological conditions, learning facilities and infrastructure, parental support, and the learning environment.

Overall, the developed instrument is capable of measuring two aspects simultaneously: misconceptions and critical thinking skills. This instrument can provide teachers with more comprehensive information regarding students' understanding levels, thus providing a basis for designing more appropriate learning and follow-up. This finding aligns with previous research, which found that integrated assessment is more effective in providing a picture of students' abilities than assessments that only measure one aspect of ability. Therefore, the developed five-tier integrated misconception and critical thinking assessment instrument can be an effective alternative assessment to support improving the quality of chemistry learning.

Conclusion

Based on the results of research and development of an integrated five-tier misconception assessment instrument for critical thinking skills on chemical bonding material, it can be concluded that: (1) The developed five-tier diagnostic assessment instrument demonstrated good quality, with satisfactory validity, very high reliability, appropriate difficulty levels, and good to very good discriminatory power. The instrument is therefore suitable for assessing students' understanding of chemical bonding. (2) The instrument was considered highly feasible and easy to use by students and teachers. Its five-tier structure provides

more detailed information about students' conceptual understanding, confidence, reasoning, and sources of information. (3) The instrument was able to identify differences in students' conceptual understanding and misconception profiles across school categories. It also successfully assessed students' critical thinking skills and detected significant differences among school categories. (4) The main contribution of this study is the development of a validated five-tier diagnostic instrument that simultaneously identifies misconceptions and assesses critical thinking skills. This integration provides a more comprehensive approach to chemistry diagnostic assessment and can support teachers in evaluating learning difficulties and designing more appropriate instructional strategies. Future studies should validate the instrument with larger and more diverse samples and extend its application to other chemistry topics.

Acknowledgments
Thank you to all parties who have helped in this research so that this article can be published.

Author Contributions
All authors contributed to writing this article.

Funding
No external funding.

Conflicts of Interest
No conflict interest.

References

Cahyana, U., Supatmi, S., Erdawati, & Rahmawati, Y. (2019). The influence of web-based learning and learning independence toward student's scientific literacy in chemistry course. *International Journal of Instruction*, *12*(4), 655–668. https://doi.org/10.29333/iji.2019.12442a

Doyan, A., Susilawati, Harjono, A., Annam, S., Muliyadi, L., & Hamidi. (2026). Development of an Interactive Web-Based STEM-PjBL Physics Material Model to Improve the Creativity of Prospective Student Teacher: A Review. *Current Educational Review*, *2*(3), 139–149. https://doi.org/10.56566/cer.v2i3.1034

Doyan, A., Susilawati, Kosim, Wardiawan, Z., Hakim, S., Muliyadi, L., & Hamidi. (2020). The development of physics module oriented generative learning to increase the cognitive learning outcomes and science process skills of the students. *Journal of Physics: Conference Series*, *1521*(2), 022059. https://doi.org/10.1088/1742-6596/1521/2/022059

Doyan, A., Susilawati, S., Annam, S., Muliyadi, L., Megahati, R. R. P., Hutabarat, R. A., Ikhsan, M., Ardianti, N. R., & Hamidi, H. (2024). Trends Research Creative Thinking Skills in Students' Physics Learning (2015-2024): A Systematic Review. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 333–342. https://doi.org/10.29303/jppipa.v10i6.7826

Ennis, R. H. (2018). Critical Thinking Across the Curriculum: A Vision. *Topoi*, *37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Hawa, A. M., & Putra, L. V. (2018). PISA Untuk Siswa Indonesia. *Janacitta*, *1*(1), 12–20. https://doi.org/10.35473/jnctt.v1i1.13

Hidayah, P., & Muchtar, Z. (2022). Pengembangan tes diagnostik berbasis web pada materi larutan elektrolit dan nonelektrolit. *Educenter : Jurnal Ilmiah Pendidikan*, *1*(6), 396–405. https://doi.org/10.55904/educenter.v1i6.190

Hidayati, U. N., & Sumarti, S. S. (2019). Desain instrumen tes three tier multiple choice untuk analisis pemahaman konsep peserta didik. *Inovasi Pendidikan Kimia*, *13*(2), 2425–2436. https://doi.org/10.15294/jipk.v13i2.19382

Kartini, K., Doyan, A., Kosim, K., Susilawati, S., Khasanah, B. U., Hakim, S., & Muliyadi, L. (2019). Analysis of Validation Development Learning Model Attainment Concept to Improve Critical Thinking Skills and Student Learning Outcomes. *Jurnal Penelitian Pendidikan IPA*, *5*(2), 185–188. https://doi.org/10.29303/jppipa.v5i2.262

Laksono, P. J. (2020). Pengembangan Three Tier Multiple Choice Test Pada Materi Kesetimbangan Kimia Mata Kuliah Kimia Dasar Lanjut. *Orbital: Jurnal Pendidikan Kimia*, *4*(1), 44–63. https://doi.org/10.19109/ojpk.v4i1.5649

Ma, C., Laeli, H., & Maret, U. S. (2021). Development of Three Tier Multiple Choice Diagnostic Test to Identify Misconception and Improve Critical Thinking Skill in Science Learning. *İlköğretim Online*, *20*(2), 277–287. https://doi.org/10.17051/ilkonline.2021.02.21

Maritsa, A., Hanifah Salsabila, U., Wafiq, M., Rahma Anindya, P., & Azhar Ma'shum, M. (2021). Pengaruh Teknologi Dalam Dunia Pendidikan. *Al-Mutharahah: Jurnal Penelitian Dan Kajian Sosial Keagamaan*, *18*(2), 91–100. https://doi.org/10.46781/al-mutharahah.v18i2.303

Muliyadi, L., Doyan, A., Susilawati, & Hamidi. (2026). The Effect of the Problem Based Learning Model Assisted by PhET Simulation on Students' Physics Learning Outcomes. *International Journal of Science Education and Science*, *3*(2), 22–26. https://doi.org/10.56566/ijses.v3i2.738

Muliyadi, L., Hamidi, H., Islami, A. V., Rizaldi, D. R., Hasanah, R., Yadaeni, A., Hariadi, M. H., Hanafi,

M., & Wathoni, H. (2026). Analysis Validation of Physics Learning Devices Based on Problem Based Learning Assisted by PhET Simulation to Improve Students' Creativity and Critical Thinking. *Indonesian Journal of Science Education and Applied Research*, *1*(1), 27–31. Retrieved from https://journals.balaipublikasi.id/index.php/ijosear/article/view/743

Prianti, T., Susanti VH, E., & Indriyanti, N. Y. (2020). Misconceptions of High School Students in Salt Hydrolysis Topic Using a Three-Tier Diagnostic Test (TTDT). *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *5*(1), 32. https://doi.org/10.20961/jkpk.v5i1.34502

Putra, A. S. U., Hamidah, I., & Nahadi. (2020). The development of five-tier diagnostic test to identify misconceptions and causes of students' misconceptions in waves and optics materials. *Journal of Physics: Conference Series*, *1521*(2). https://doi.org/10.1088/1742-6596/1521/2/022020

Rokhim, D. A., Rahayu, S., & Dasna, I. W. (2023). Analisis Miskonsepsi Kimia dan Instrumen Diagnosisnya: Literatur Review. *Jurnal Inovasi Pendidikan Kimia*, *17*(1), 17–28. https://doi.org/10.15294/jipk.v17i1.34245

Rosita, I., Liliawati, W., & Samsudin, A. (2020). Pengembangan Instrumen Five-Tier NewtonÃ¢â¬â¢s Laws Test (5TNLT) Untuk Mengidentifikasi Miskonsepsi dan Penyebab Miskonsepsi Siswa. *Jurnal Pendidikan Fisika Dan Teknologi*, *6*(2), 297–306. https://doi.org/10.29303/jpft.v6i2.2018

Rosyana, W., Ashadi, A., & Mulyani, S. (2019). Pengembangan Instrumen Penilaian Three-Tier Multiple Choice (Three- TMC) Untuk Mengukur Keterampilan Berpikir Kritis Siswa Pada Materi Kimia Larutan Penyangga Kelas XI SMA. *INKUIRI: Jurnal Pendidikan IPA*, *8*(1), 48. https://doi.org/10.20961/inkuiri.v8i1.31814

Rumapea, C. F., & Silaban, R. (2022). Pengembangan Instrumen Tes Diagnostik Three-Tier Multiple Choice Berbasis Android Based Test Untuk Mengukur Miskonsepsi Siswa Pada Materi Laju Reaksi Kelas XI SMA. *Educenter : Jurnal Ilmiah Pendidikan*, *1*(2), 95–104. https://doi.org/10.55904/educenter.v1i2.44

Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., Hu, S., Yu, N., Ma, L., Guo, F., Su, Q., Shi, J., Shi, H., & Loyalka, P. (2020). Thinking critically about critical thinking: validating the Russian HEIghten® critical thinking assessment. *Studies in Higher Education*, *45*(9), 1933–1948. https://doi.org/10.1080/03075079.2019.1672640

Suparno, P. (2013). *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. PT Grasindo.

Suwarto. (2013). *Pengembangan Tes Diagnostik dalam Pembelajaran*. Pustaka Pelajar.

Teknowijoyo, F. (2022). Relevansi Industri 4.0 dan Society 5.0 Terhadap Pendidikan Di Indonesia. *Educatio*, *16*(2), 173–184. https://doi.org/10.29408/edc.v16i2.4492

Widiandari, L. A., & Redhana, I. W. (2021). Students' critical thinking skills in case study-based learning. *AIP Conference Proceedings*, *2330*(March). https://doi.org/10.1063/5.0043204

Zakiah, L., & Lestari, I. (2019). *Berpikir Kritis Dalam Konteks Pembelajaran*. Erzatama Karya Abadi.