E-ISSN: 2746-7767
Ainara Journal (Jurnal Penelitian dan PKM Bidang Ilmu Pendidikan)
Volume 7, Issue 3, September 2026, pp 502-517
https://doi.org/10.54371/ainj.v7i3.1456

Development of A Science Literacy-Based Four-Tier Diagnostic Instrument to Identify Misconceptions in Chemical Bonding Material

***Junia Marsanita[1], Wiwi Siswaningsih[2], Febriati Dian Mubarokah[3]**
[1,2,3]Universitas Pendidikan Indonesia, Indonesia
Correspondence Email: juniamarsanita14@gmail.com

Article History: Submission: 2026-05-10 || Accepted: 2026-09-01 || Published: 2026-09-20

Abstract

This study aimed to develop a science-literacy-based four-tier diagnostic instrument for identifying students' misconceptions in chemical bonding and to establish evidence of its content validity. The instrument was developed using the Test Design and Development approach, focusing on the planning and development stages. An initial pool of 40 multiple-choice items integrated into nine contextual informational texts was evaluated by five expert validators, comprising three chemistry education lecturers and two high school chemistry teachers. Content validity was assessed using the Content Validity Ratio (CVR), Content Validity Index (CVI), Item-Level Content Validity Index (I-CVI), and Scale-Level Content Validity Index based on the average method (S-CVI/Ave). The results showed that 37 of the 40 items obtained a CVR of 1 and met the required criterion for five validators, while three items failed to satisfy the threshold and therefore required further revision or exclusion before pilot testing. Based on the I-CVI analysis, all 40 items met the acceptable criterion of ≥ 0.78. At the scale level, the instrument achieved an overall CVI of 0.9875 and an S-CVI/Ave of 0.99375, indicating a high level of agreement among experts regarding the relevance of the instrument content. These findings provide strong preliminary evidence for the content validity of the developed instrument but do not establish its overall psychometric quality or diagnostic accuracy. Further empirical testing involving students is therefore required to examine construct validity, reliability, item functioning, classification accuracy, and diagnostic performance before the instrument can be recommended for broader diagnostic use.

Keywords: *Chemical bonding, Content validity, Four-tier diagnostic test, Misconceptions, Science literacy.*

I. INTRODUCTION

21st-century learning emphasizes mastery of four key skills known as the 4Cs (creative thinking, critical thinking, communication, and collaboration). These skills serve as the foundation for students to adapt to rapid changes in the workplace (Septikasari & Frasandy, 2018). Furthermore, the approach to science education in this era has shifted from teacher-centered learning to student-centered learning, where students are encouraged to learn independently, explore, and apply knowledge in real-world contexts, particularly in science literacy (Juliartawan & Riastini, 2022; Pradini, Wijaya, & Jannah, 2022; Yuliati, 2017). This approach aims to develop students who not only memorize facts but are also capable of analyzing, implementing, and solving science-based problems.

The Organisation for Economic Cooperation and Development (OECD) defines science literacy as an individual's ability to effectively apply scientific knowledge to identify scientific problems, build new understandings, formulate explanations based on empirical evidence, draw conclusions grounded in scientific principles, and develop a reflective attitude to participate in scientific discussions and problem-solving within society (Rohmaya, 2022; OECD, 2023). This definition emphasizes the dimensions of application and context, not merely mastery of subject matter.

Students' science literacy skills worldwide are periodically assessed through the Programme for International Student Assessment (PISA), conducted every three years by the OECD. PISA results serve as a reflection of a country's educational quality. Unfortunately, PISA data indicate that Indonesian students' science literacy performance remains far below the international average. Over nearly 20 years of participation (since 2000), Indonesia has consistently ranked among the bottom 10 in the science literacy category (Rohmaya, 2022). In the 2022 PISA cycle, although Indonesia's ranking improved by 6 places from 71st to 65th among 81 countries (compared to 2018) its absolute score actually dropped by 13 points (from 396 to 383), which is nearly in line with the 12-point decline in the international average (Kemendikbudristek, 2024; OECD, 2023).

Table 1. Science Literacy Skills of Indonesian Students Based on PISA Results from 2000 to 2018 (OECD)

Year	Indonesia's Average Score	International Average Score	Number of Participating Contries	Indonesia's Ranking
2000	393	500	41	38
2003	395	500	40	38
2006	393	500	56	50
2009	383	500	65	60
2012	382	500	65	64
2015	403	500	69	62
2018	396	500	79	71

To address the crisis in education quality, the Indonesian government has implemented the Merdeka Curriculum and priority programs as key innovations. According to Nahadi et al. (2022), this curriculum is designed to create an ideal, flexible, and enjoyable learning environment, with a focus on developing students' individual potential. The Merdeka Curriculum integrates four key elements: content knowledge, literacy (reading, numeracy, science), 21st-century skills (4Cs), and a positive attitude toward the use of digital technology (Nahadi et al., 2022). The Merdeka Curriculum is well-suited to developing 21st-century skills through science-literacy-based learning. For example, project-based learning can be used to integrate real-world science contexts, such as environmental impact analysis or technological innovation. However, the success of this curriculum depends on teachers' ability to identify students' difficulties, including misconceptions that often serve as major barriers (Muliaman et al., 2022).

Curriculum changes have been implemented, yet science education still faces the challenge of misconceptions. That is, erroneous concepts or alternative understandings held firmly by students and difficult to change through conventional teaching methods (Treagust, 1988). Misconceptions arise from various sources such as everyday intuitive experiences (for example, that "water is a neutral substance"), the influence of everyday language, flawed teaching models, or a lack of understanding of basic concepts (Sudiatmika et al., 2025; Yuberti et al., 2020; Treagust, 1988). In chemistry, these misconceptions are particularly prominent in the topic of chemical bonding, an abstract concept that serves as the foundation for understanding atomic and molecular structures as well as chemical reactions (Islami et al., 2019).

Research by Islami et al. (2019) indicates that high school students in Indonesia exhibit a misconception rate of up to 30–31% regarding chemical bonding. Significant misconceptions were identified in the subtopics of Lewis structures and the octet rule (33.33%), metallic bonding and metallic properties (20.83%), ionic and covalent compounds (27.08%), VSEPR theory (20.83%), electron domain theory (18.75%), molecular polarity (27.08%), van der Waals forces (14.58%), and hydrogen bonding (29.17%). For example, viewing ionic bonds as "electron exchange" rather than electron transfer, difficulty distinguishing between polar and nonpolar covalent bonds, and misunderstanding hydrogen bonding as a "strong chemical bond" (Hendrawani, 2023; Rahayu et al., 2021; Yasthopi & Ritonga, 2017). These misconceptions not only lower academic achievement but also hinder science literacy because students fail to apply concepts in future contexts.

To accurately identify misconceptions, diagnostic instruments are essential. The two-tier diagnostic instrument (answer-reason selection) has evolved into a four-tier model, consisting of: (1) conceptual statements, (2) confidence level in the answer, (3) reasoning, and (4) confidence level in the reasoning (Sudiatmika et al., 2025; Yuberti et al., 2020; Treagust, 1988). This model can classify student responses into three categories: conceptual understanding, lack of conceptual understanding, and misconceptions (Sudiatmika et al., 2025; Yuberti et al., 2020). Many four-tier diagnostic instruments have been developed, but none have been developed using a science literacy approach. Yet integrating science literacy learning with the PISA framework (contextualization, reflection, and application) can enhance the effectiveness and conceptual understanding of students. If this four-tier diagnostic instrument is combined with contextual science literacy-based learning through the PISA framework which emphasizes the ability to explain scientific phenomena, evaluate information, and apply concepts in daily life the reduction in misconceptions is likely to be more significant. A learning approach that links chemical bonding concepts to real-world phenomena such as the mechanism of salt dissolution, iron corrosion, metal characteristics, and the boiling points of compounds can help students develop a deeper conceptual understanding rather than relying solely on memorizing theoretical formulas.

Accordingly, this study aims to develop a science-literacy-based four-tier diagnostic instrument for identifying misconceptions in chemical bonding and to evaluate its content validity through expert judgment. At this stage of development, the study focuses specifically on evidence of content validity as a prerequisite for subsequent empirical testing of the instrument with students.

II. METHOD

This study employed the Test Design and Development approach, which is appropriate for developing and validating assessment instruments (ERA, APA, & NCME, 2014; Adams & Wieman, 2011). The framework comprises four stages: (1) planning, (2) development, (3) field testing, and (4) evaluation for operational use. Given the scope of the present study, the research focused on the first two stages—planning and development—to develop a science-literacy-based *four-tier* diagnostic instrument for identifying misconceptions in chemical bonding. The detailed procedures undertaken in these stages are presented in Figure 1.

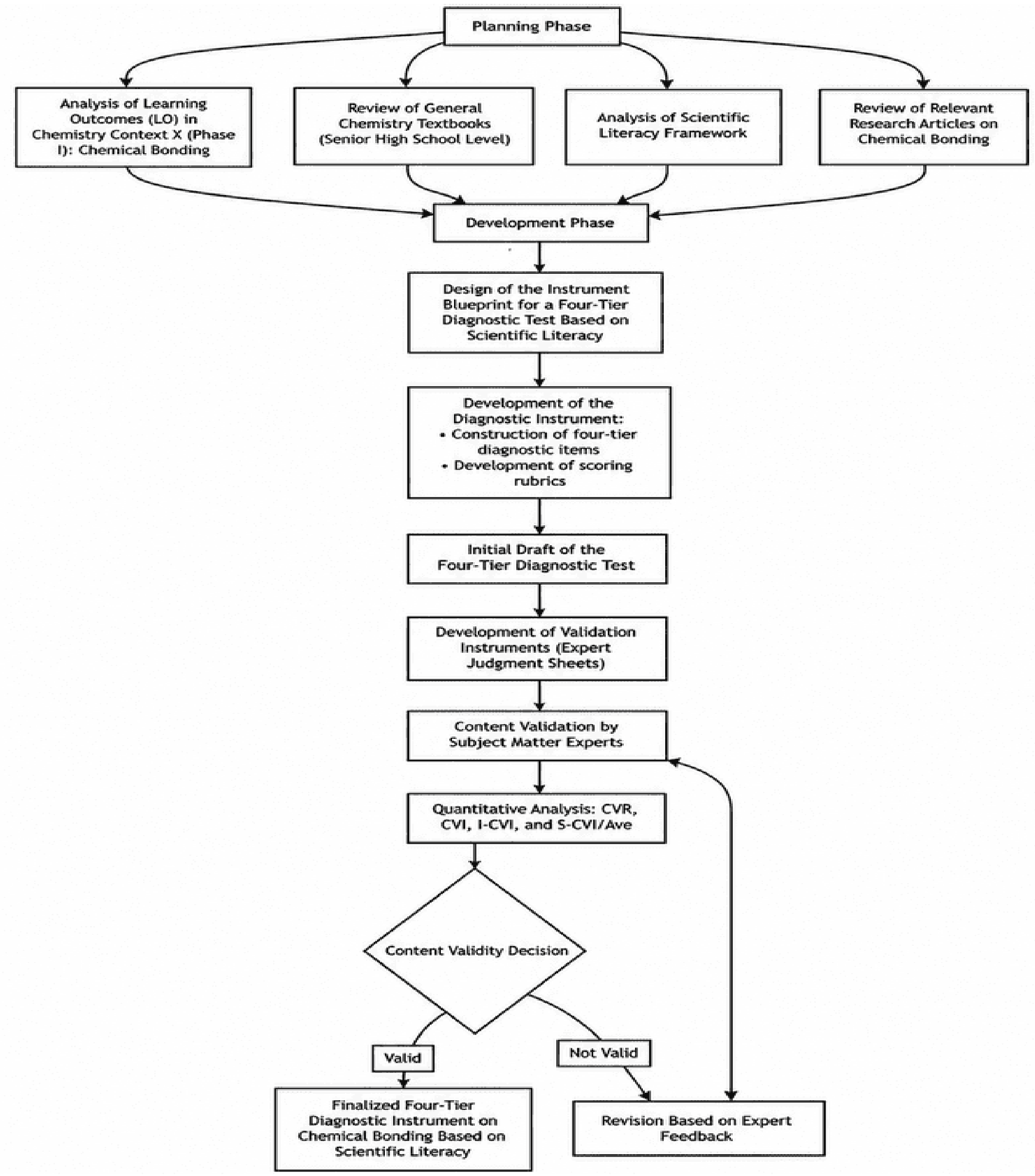

Figure 1. Instrument Development Process Flowchart

The research participants who served as validators to evaluate the validity of the designed assessment instrument consisted of five individuals: three Chemistry Education lecturers from the Faculty of Mathematics and Natural Sciences (FPMIPA) at UPI and two high school chemistry teachers in Bandung.

The research instruments used included an item matrix, draft test items, and a validation sheet. The item matrix serves as a basic guideline or framework for designing *four-tier* diagnostic questions, which will be formulated by the researcher and answered by students. The developed item matrix includes elements such as learning outcomes, subtopics, aspects of science literacy, misconceptions to be identified, item indicators, question formats, and item numbers. The validation sheet is used to check the alignment of test items with the characteristics of the four-tier diagnostic instrument, the alignment of test items with aspects of science literacy, the alignment of item indicators with the test items, as well as the clarity and accuracy of answer options in identifying misconceptions. The validation table includes columns for expert input regarding corrections if errors or inconsistencies are found, which are described narratively. Experts involved in the validation were provided with a special form and asked to complete it by checking (√) the appropriate columns and including comments, recommendations, or suggestions for improvement.

Data analysis was conducted quantitatively based on the results of the expert validation. To objectively assess the validity of the validation results for the four-tier science literacy-based diagnostic instrument, data were analyzed using the Content Validity Ratio (CVR) for each item aspect, as well as the Content Validity Index (CVI), which includes the Item-Content Validity Index (I-CVI) and the Scale-Content Validity Index/Average (S-CVI/Ave) for the entire instrument. These values were compared with standard categories to determine the instrument's suitability for the pilot testing phase or for further refinement.

According to Lawshe (1975), the following steps are used to obtain content validity data: a *four-tier* science literacy-based diagnostic instrument is validated by validators. Content validity scores are calculated using the *Content Validity Ratio* (CVR) with Lawshe's formula. The CVR result is then compared with the minimum CVR value for those five validators. The measurement tool is considered valid if the CVR value obtained in this study is greater than 0.99, as shown in Table 2 below:

Table 2. Minimum CVR Values (Lawshe, 1975)

Number of Validators	Min. CVR Value	Number of Validators	Min. CVR Value
5	0,99	13	0,54
6	0,99	14	0,51
7	0,99	15	0,49
8	0,75	20	0,42
9	0,78	25	0,37
10	0,62	30	0,33
11	0,59	35	0,31
12	0,59	40	0,29

CVR is used to determine whether a test item should be retained or rejected. Once the test items to be included in the final version have been identified, the next step is to calculate the content validity index (CVI) to assess the overall validity of the instrument. According to Lawshe, the CVI is derived from the average CVR score of the retained test items. Lawshe (1975) did not establish a specific threshold for the CVI as an indicator of perfect validity, and according to him, there is no standard interpretation of CVI values. The CVI merely provides an indication of the level of agreement among experts (validators) regarding overall content validity. To obtain more in-depth CVI values, including the *Item-Content Validity Index* (I-CVI) and the *Scale-Content Validity Index Average* (S-CVI/Ave), the procedure followed in this study was that used by Polit et al. (2007), as follows:

a. The validators' validation results were categorized on a scale of 1–4, as follows: the criterion "valid" was scored as 4, "valid with revisions" as 3, and "invalid" as 1 or 2. For each item, the I-CVI value was obtained by counting the number of validators who assigned a score of 3 or 4, then dividing that number by the total number of validators.
b. To obtain the S-CVI/Ave value, the total I-CVI score for all items is calculated and then divided by the total number of items. An I-CVI score of ≥ 0.78 and an S-CVI/Ave score of at least ≥ 0.90 indicate that all item scales in the instrument have excellent content validity (Polit *et al.*, 2007).

III. RESULTS AND DISCUSSION

A. Results

1. Planning Phase

The initial planning phase began with an analysis of the Learning Outcomes (LO) and Learning Objectives (LO) for Phase F Chemistry as outlined in Decision No. 046/H/KR/2025 of the Head of the Agency for Educational Standards, Curriculum, and Assessment, Ministry of Education, Culture, Research, and Technology. This analysis aims to determine the core content, relevant scope, as well as the breadth and depth of the material, so that it can be aligned with the developed instrument, namely a four-tier science literacy-based diagnostic instrument focused on identifying misconceptions. Based on the analysis, it was determined that chemical bonding material is the primary focus. This material was then broken down into several subtopics, including: 1) the duplet and octet rules, 2) Lewis dot symbols, 3) ionic bonds, 4) covalent bonds, 5) metallic bonding, 6) molecular shape, 7) intermolecular forces, 8) the influence of intermolecular forces on the physical properties of substances, and 9) the application of chemical bonding and intermolecular force concepts.

Within the PISA 2025 framework, science literacy encompasses three aspects: context, knowledge, and competence, which form the basis for instrument development. Since chemical bonding is a fundamental concept in chemistry relevant to real-world applications, the contexts used focus on six main areas: 1) health technology, 2) renewable energy production, 3) technological processes, 4) environmental impacts of mining, 5) living organisms, and 6) environmental impacts, to ensure the relevance of these contexts to daily life. Content knowledge emphasizes the understanding of key scientific concepts from well-established, up-to-date, and enduring theories of chemical bonding, such as covalent, ionic, and metallic bonds in a molecular context. Science literacy competencies encompass two processes: 1) explaining phenomena scientifically by recalling and applying appropriate scientific knowledge and identifying, constructing, and evaluating models (e.g., Lewis or VSEPR models for chemical bonding), as well as 2) investigating, evaluating, and using scientific information for decision-making and action, with related sub-competencies requiring students to analyze experimental data or simulations of chemical bonding in the surrounding environment.

Furthermore, an analysis was conducted on high school chemistry textbooks and General Chemistry textbooks to ensure that the diagnostic instruments developed align with correct chemical concepts and are relevant to learning standards. Textbooks, whether for students or teachers, play a role in identifying the core material used in instrument development, adjusting the difficulty level of questions to students' abilities, and ensuring that the developed diagnostic instruments have a strong scientific foundation. Thus, the results of this textbook analysis serve as a crucial foundation for maintaining the accuracy of the material used to understand the characteristics of each test item, ensuring that instrument development aligns with key components of science literacy tailored to chemical bonding content.

Subsequently, a literature review was conducted on various scientific articles and journals regarding misconceptions in chemical bonding material. From these misconceptions, concepts that should be correct were also identified based on previous analyses of the textbook and General Chemistry. These misconceptions will later be used as answer options (distractors) to detect misconceptions and reinforce conceptual understanding based on curriculum standards, thereby improving the accuracy of the four-tier science literacy-based diagnostic instrument being developed.

In line with this, the development of the four-tier science literacy-based diagnostic instrument also refers to the results of a literature review by Yuberti et al. (2020) regarding the characteristics of a four-tier diagnostic instrument consisting of four levels, namely: 1) the first tier consists of a conceptual statement with four answer choices (one correct, three misconception distractors), 2) the second tier measures the level of confidence in the answer (confident or not confident), 3) the third tier asks for the reason for selecting the answer with options for scientific explanations, and 4) the fourth tier measures the level of confidence in the reason (confident or not confident). This framework guides the development of the instrument to distinguish among three categories of student understanding (conceptual understanding, lack of conceptual understanding, and misconceptions). Thus, this entire planning process ensures that the developed instrument is not only aligned with the curriculum and relevant literature but is also capable of identifying students' misconceptions.

2. Development Phase

The development phase included activities such as designing the framework for the four-tier diagnostic instrument, drafting informational texts, developing items for the four-tier diagnostic test, and designing assessment rubrics, which culminated in the preparation of an initial draft of the science literacy-based four-tier diagnostic instrument. During this stage, validation is also conducted through content validity assessment by five expert validators using a pre-prepared validation sheet. If the instrument is deemed valid, the research can proceed to the next stage. Conversely, if the instrument does not yet meet the validity criteria, revisions or refinements are made until a draft instrument suitable for pilot testing is obtained. This development stage is the second step in the research process, specifically focused on the development of a four-tier diagnostic instrument based on science literacy. The first step in this stage is to develop a blueprint for the four-tier science literacy-based diagnostic instrument, which serves as a guide for developing the items of the four-tier science literacy-based diagnostic test. This blueprint is developed by referring to the PISA 2025 framework, ensuring that each test item indicator truly reflects the desired competencies.

This instrument's item bank is structured and includes various key components, such as the title of the contextual reading text or informative text, learning outcomes (LO) for the chemistry subject being studied, the subtopic of chemical bonding, aspects of science literacy according to the PISA framework, misconceptions to be identified, item indicators, question format, and item number. The development of this item matrix aims to ensure that the items created have a strong foundation and align with the research objectives. For example, Table 3. presents an example of a four-tier science literacy-based diagnostic instrument matrix designed. The presentation of this table aims to provide a concrete illustration of the application of the principles of developing a four-tier science literacy-based diagnostic instrument.

Table 3. Outline of a four-tier diagnostic instrument based on science literacy

LO	Sub-Topic	Aspects of Science Literacy	Miscon ception	Item Indicator	Ques tion Format	Ques tion Number
Comparing types of chemical bonds and their relationship to molecular shape and intermolecular forces in predicting the physical properties of matter	Covale nt Bond	PISA 2025 – Knowledge Aspect (Content Knowledge): Relevant to real-life situations; represents key scientific concepts derived from well-established and up-to-date theories that remain useful over the long term. Competence Aspect (Explaining Phenomena Scientifically): Recalling and applying appropriate scientific knowledge. Context Aspect: Health and Technology.	A bond formed by the sharing of a pair of electrons between two atoms.	Identify the type of covalent bond in the O_2 molecule (double bond).	Multiple -choice	15
Same as above	Covale nt Bond	Same as above	A bond formed by the sharing of two electrons between two atoms.	Identify the type of covalent bond in the O_2 molecule (double bond).	Multiple -choice	15
Same as above	Covale nt Bond	Same as above	Bonds are formed by atoms with	Identify the type of covalent	Multiple -choice	15

	a valence of 2 or 6.	bond in the O_2 molecule (double bond).	

Once the item matrix development process is complete, the next stage is the development of the four-tier science literacy-based diagnostic instrument. This process begins with the creation of contextual readings or informative texts relevant to the core material on chemical bonding. These texts are structured as question stimuli, intended to illustrate the application of chemical bonding concepts in various everyday contexts. The text as a whole is presented through paraphrasing using clear scientific language appropriate to students' comprehension levels, facilitating the use of concepts such as covalent bonding, molecular polarity, intermolecular interactions, and compound stability through authentic applications. Therefore, this text serves as an introduction to a four-tier science literacy-based diagnostic instrument, linking the principles of chemical bonding with innovations in renewable technology.

After the initial draft of the four-tier science literacy-based diagnostic instrument was completed, the next step was to evaluate its quality through content validity testing. This validation process aims to measure the extent to which the instrument's content aligns with the competencies intended for assessment. This step is crucial because it ensures that each item accurately reflects the intended competencies without deviating from the research objectives (Juliani & Erita, 2023). A valid instrument is capable of accurately describing students' abilities or performance in alignment with the formulated learning objectives (Yusup, 2018). Beyond mere measurement accuracy, validation also aims to identify shortcomings in the instrument, including aspects of content, item formulation, and alignment with indicators. Validation results serve as the basis for revisions to produce high-quality, functional instruments that meet evaluation needs (Dewi et al., 2020). This study employs content validity, which involves assessing appropriateness through expert judgment from specialists in the relevant field (Firman, 2013; Puspitasari & Febrinita, 2021). Content validity evaluates the extent to which test items cover subject matter domains aligned with learning outcomes, both in terms of substance and cognitive demands (Firman, 2013). An optimal level of content validity indicates that the instrument has met competencies in a representative manner (Sensus et al., 2022). Thus, in the development of a four-tier science literacy-based diagnostic instrument, content validity not only ensures the accuracy of the content but also detects deficiencies before pilot testing.

Content validity describes the extent to which the design of a measurement instrument aligns with the learning material being evaluated. An instrument is considered to have adequate content validity if each item effectively measures learning objectives that are consistent with the material that has been taught (Arikunto, 2016; Ramadhan et al., 2024). Content validity assessment is generally conducted through an analysis of the test blueprint, which includes the material and the targeted cognitive levels, and this process is carried out before the instrument is administered (Mukhlisa, 2023). Good content validity indicates that the test items in the instrument comprehensively represent the entire subject matter, including the content aspects and critical thinking skills required to answer them. The content validity evaluation process is carried out through assessment by experts competent in the relevant field (Firman, 2013).

In this study, content validity was assessed by reviewing and evaluating the draft of the previously designed science literacy-based four-tier diagnostic instrument. This stage involved five validators: three chemistry education lecturers and two teachers from a high school. The validators were asked to assess the alignment of the test items with the characteristics of the four-tier diagnostic instrument, the alignment of the test items with aspects of science literacy, the alignment of the item indicators with the test items, as well as the clarity and accuracy of the answer options in identifying misconceptions. During the content validation of the science literacy-based four-tier diagnostic instrument, validators marked the "Yes" column with a checkmark (√) if the aspect was deemed appropriate, or the "No" column if it was deemed inappropriate. Additionally, a "Suggestions" column was provided to list recommendations for improvements or feedback regarding the reading text, item indicators, item wording, and scoring rubric in the designated columns.

According to Lawshe (1975), the content validity of a measurement instrument can be evaluated through expert judgment using the dichotomy method, in which each test item is categorized as "essential" or "non-essential" in relation to the competency being measured. Items considered essential by most experts are deemed to represent the appropriate construct. In this study, a similar method was applied using "Yes" or "No" choices to assess the alignment of test items with criteria such as reading passages and their indicators. Lashe also introduced the Content Validity Ratio (CVR) formula as a quantitative tool to measure the level of expert consensus regarding the essentiality of an item, calculated as the proportion of experts who agree and compared to a critical threshold based on the number of validators. For example, with five validators, a minimum CVR value of 0.99 is required for the item to be considered valid and worthy of retention.

In addition to assessing validity at the item level, the Content Validity Index (CVI) is used to evaluate overall content validity by calculating the average CVR of items that meet the criteria. Although Lawshe does not explicitly equate the "essential" category with "valid," the assessment of essentiality by most experts implicitly indicates the content's relevance to the measurement domain, thereby serving as an indicator of content validity. This approach has been widely applied in studies, such as Suryadi et al. (2023) and Puspitasari & Febrinita (2021), which utilize CVR, CVI, I-CVI, and S-CVI/Ave as quantitative measures of item or instrument validity, provided their values meet the threshold.

The results of the content validity assessment reflect the validators' evaluation of four main aspects: the alignment of test items with the characteristics of the four-tier diagnostic instrument, the alignment of test items with aspects of science literacy, the alignment of item indicators with the test items, and the clarity and accuracy of response options in identifying misconceptions. The CVR scores for each aspect are presented in Table 4.

Table 4. Overall CVR

Content Validity Categories	The Number of Valid Items Based on The CVR	Description
Alignment of test items with the characteristics of a four-tier diagnostic instrument	39	Question 10 is invalid
Alignment of test items with aspects of science literacy	39	Question 10 is invalid
Alignment of item indicators with test items	38	Question 10 and 22 are invalid
Clarity and accuracy of response options in identifying misconceptions	39	Question 9 is invalid

The calculated CVR values are shown in Table 4. These are then compared with the minimum CVR threshold values according to Lawshe's (1975) criteria, adjusted for the number of validators. For a group consisting of five validators, the minimum CVR value is 0.99. An item is considered valid and worth retaining if the obtained CVR value exceeds 0.99, whereas the item is considered invalid and must be discarded if its CVR value is less than 0.99.

37 test items were deemed valid because their CVR values exceeded 0.99 specifically reaching 1, so these items are considered suitable to proceed to the pilot testing phase with students. These items have met content validity standards according to experts, including alignment with competency indicators, clarity of the stimulus, clarity of answer options, and precision in item formulation. There are several test items that can still be used, although they require adjustments to certain elements based on suggestions from the validators; thus, they can be used in the pilot test after being revised in accordance with the experts' recommendations. Conversely, there are 3 test items that are invalid because their CVR values are below 0.99, so they will not be included in the next pilot test phase.

Item number 9 was deemed invalid based on content validity, specifically regarding the clarity and accuracy of the answer options in identifying misconceptions in accordance with the characteristics of a four-tier science literacy-based diagnostic instrument, where the distractor options in tiers 1 and 3 must represent the misconceptions intended to be revealed in order to distinguish conceptual understanding from mere guessing. In Tier 1, which asks about VSEPR notation and the shape of H_2O, distractor options such as linear/tetrahedral shapes are too easy to rule out because the bent shape is already known, thus failing to force in-depth analysis. In Tier 3, the justifications for the distractor options are too lengthy, formulaic, explicitly incorrect (e.g., molecular shape is not influenced by lone pairs), or too theoretical (such as orbital hybridization), thus failing to reflect common misconceptions like equating molecular shape with electron domains or the assumption that molecules bonded by two atoms are always linear. Consequently, this item is ineffective as a diagnostic tool for identifying patterns of misconceptions regarding VSEPR theory and molecular shape.

Item number 10 is also considered invalid based on content validity because it does not fully align with the characteristics of a four-tier diagnostic instrument, aspects of science literacy, indicator appropriateness, and the quality of the answer options used. The context of boiling point graphs for hydrogen compounds should require an analysis of the relationship between relative molecular mass (Mr), intermolecular forces, and hydrogen bonding in compounds such as H_2O, HF, and NH_3. However, the item only emphasizes the classification of polar/nonpolar covalent bonds, thus failing to measure graphical interpretation and explanation of boiling point anomalies. Tier 1 and Tier 3 are not sufficiently in-depth in analyzing misconceptions, with weak inter-level connections and ineffective answer options (e.g., the option regarding ionic/coordination bonds is easily refuted, and general justifications do not represent actual misconceptions, such as the assumption that high electronegativity always results in the highest boiling point). Consequently, the item does not optimally cover science literacy indicators and misconception analysis, thus requiring a comprehensive revision before being used for pilot testing.

Item number 22 is considered invalid due to a mismatch between the item's indicators and the item itself regarding the formation of the ionic compound NaCl. The indicators are intended to assess the ability to explain the ionic bonding process through the transfer of electrons from Na to Cl to achieve octet stability. However, Tier 1 focuses on memorizing the initial electron configurations of the Na and Cl atoms based on their atomic numbers. This allows for a correct answer without analyzing the formation of the ionic compound. Meanwhile, Tier 3, which addresses stability, is not synchronized with Tier 1, thereby weakening the diagnostic ability to identify misconceptions. The question emphasizes the symbolic representation of electron configurations rather than the process of NaCl compound formation, requiring a comprehensive revision to align with the item's indicator.

Next, after calculating the CVR values, the Content Validity Index (CVI) was calculated, which is the average CVR of the items retained and deemed suitable for pilot testing (Lawshe, 1975). The CVI value serves to describe the overall level of content validity of the developed instrument. Details of the CVI values based on content validity categories and the total CVI value are presented in Table 5. A high CVI value indicates that the developed instrument has adequate content appropriateness and is ready to proceed to the next stage in the development of a four-tier science literacy-based diagnostic instrument.

Table 5. Total CVI Score

Content Validity Categories	CVI Value for Each Category	Total CVI Score
Alignment of test items with the characteristics of a four-tier diagnostic instrument	0,99	0,9875
Alignment of test items with aspects of science literacy	0,99	
Alignment of item indicators with test items	0,98	
Clarity and accuracy of response options in identifying misconceptions	0,99	

Based on Table 5, the overall CVI value for the four-tier science literacy-based diagnostic instrument reached 0.9875, indicating a very high level of agreement among experts regarding the content representation of the items in relation to the measured indicators and competencies. According to Lawshe (1975), the CVI is not a correlation coefficient with a fixed interpretive threshold for validity, but rather a measure of expert agreement regarding the instrument's content representation within the measurement domain. The higher the value, the stronger the evidence that the items are suitable for measuring the targeted competencies, thus serving as a quantitative indicator of content representation without being the sole determinant of validity. In the development of this four-tier science literacy-based diagnostic instrument, a CVI value approaching 1 confirms that the majority of test items are considered appropriate and representative by experts for identifying misconceptions.

To clarify the interpretation of the CVI, the I-CVI and S-CVI/Ave were also calculated according to Polit et al. (2007), in order to clarify these indices and address ambiguities in content validity results in instrument development studies, where the S-CVI represents overall validity based on the average of the I-CVIs. The I-CVI value is derived from the proportion of experts who rated an item as relevant using a 4-point scale (1 = not relevant, 2 = somewhat not relevant, 3 = somewhat relevant, 4 = very relevant), with a score of 4 for the "valid" aspect, 3 for "valid with revision," and 1–2 for "invalid." The formula divides the number of experts who scored 3–4 by the total number of experts. Polit et al. (2007) recommend a 4-point scale rather than a 3- or 5-point scale to avoid neutral options that obscure the assessment of clarity relevance (as suggested by Lynn, 1986).

The analysis results indicate that 40 items regarding item-instrument fit for the four-tier diagnostic instrument, item-science literacy fit, indicator-item fit, and the clarity and precision of response options in identifying misconceptions were deemed valid because their I-CVI values were > 0.78. However, there were differences in content validity findings between the CVR method (Lawshe, 1975) and I-CVI. The CVR analysis identified 3 invalid items (CVR < 0.99 for five validators), whereas I-CVI indicated that all items were valid because they met the minimum criteria of Polit et al. (2007). Additionally, after calculating the I-CVI for each aspect, the S-CVI/Ave was calculated as the average of all I-CVI values, reflecting the validators' agreement regarding the overall relevance of the items and providing an indication of the scale's general content validity. This calculation is crucial for verifying the consistency of validity not only at the item level but also at the overall instrument level. Details of the S-CVI/Ave for each content validity aspect category and for the instrument as a whole are presented in Table 6.

Table 6. Overall S-CVI/Ave Score

Content Validity Categories	S-CVI/Ave Score for Each Category	Total S-CVI/Ave Score
Alignment of test items with the characteristics of a four-tier diagnostic instrument	0,995	0,99375
Alignment of test items with aspects of science literacy	0,995	
Alignment of item indicators with test items	0,99	
Clarity and accuracy of response options in identifying misconceptions	0,995	

Table 6. presents the total S-CVI/Ave values for the instrument based on content validity categories, with the categories of item alignment with the characteristics of the four-tier diagnostic instrument, item alignment with aspects of science literacy, and the clarity and accuracy of response options in revealing misconceptions achieving a score of 0.995, and the alignment of item indicators with the items achieving a score of 0.99, resulting in an average of 0.99375. According to Polit et al. (2007), an S-CVI/Ave value ≥ 0.90 indicates excellent category content validity, so the instrument as a whole is considered valid in terms of content. However, this validation process is influenced by various factors as explained by Nahadi & Firman (2019), including: 1) test characteristics such as ambiguous instructions, complex sentence structures, inconsistent item difficulty levels, poor construction, items that are not aligned with learning outcomes, time constraints, or a minimal number of items; 2) alignment between content and the sequence of learning objectives; 3) technical aspects of administration and scoring; 4) student responses; and 5) group characteristics and scoring criteria. Therefore,

although most items are valid, the development of this instrument still requires revision to address its weaknesses.

Based on the theoretical framework and validation results, the four-tier science literacy-based diagnostic instrument still has shortcomings in aspects such as the clarity of the reading text structure, the consistency of question instructions, the appropriateness of answer options, the accuracy of scoring, and the representation of competency indicators. Therefore, the revision process is crucial for improving quality. Revisions were made based on feedback from validators, including improvements to question format, wording, data completeness, sentence order, typo corrections, as well as the addition of information to the text, and refinement of image/source captions. Assessment revisions focused on adjusting questions, indicators, and weights. Examples of text and question revisions are shown in Table 7.

Table 7. Example of Improvements to a Four-Tier Science Literacy-Based Diagnostic Instrument

Instrument Components	Revised Aspects	Revised Version
Text 1	1. Correction of spelling errors in the stimulus title 2. Clarity of the illustrations in the stimulus 3. Level of information diclosure in the stimulus	1. The word “wasah” has been revised to “wadah” to improve linguistic accuracy 2. Labels A and B have been added to the images, and the image captions have been refined to make the illustrations more informative 3. The observation results have been removed so that students can analyze the phenomenon independently based on the provided stimuli
Question 8	1. Effectiveness of the item stem 2. Alignment of indicators with test items 3. Balance in the length of answer options 4. Accuracy of the representation of chemical bonding concepts	1. The question stems have been revised to be more concise and directly address the concepts being assessed 2. The wording of the questions has been adjusted to align with the learning indicators being measured 3. The distractors have been revised to ensure the length of the options is relatively balanced and does not lead students toward a spesific answer 4. The notation for the molecular formulas of oxygen and water hase been clarified to oxygen (O2) and water (H2O) to improve scientific accuracy

B. Discussion

The results of the content validity test indicate that the developed four-tier science literacy-based diagnostic instrument has a high level of content validity. This is demonstrated by the CVR, CVI, I-CVI, and S-CVI/Ave scores, with the majority of test items meeting the established validity criteria. The high level of agreement among validators indicates that the developed test items are deemed capable of representing competencies, aspects of science literacy, and misconceptions regarding chemical bonding material. According to Lawshe (1975), a high CVR value indicates strong agreement among experts regarding the essentiality of an item in measuring the intended construct. Furthermore, Polit et al. (2007) explain that an S-CVI/Ave value above 0.90 indicates excellent content validity. Thus, the validity results obtained in this study indicate that the developed instrument is suitable for identifying misconceptions regarding the concept of chemical bonding.

The high validity results are likely influenced by the systematic instrument development process carried out during the planning and development stages. The integration of curriculum analysis, textbook analysis, misconception analysis, and the PISA 2025 science literacy framework ensures that the developed test items align between competencies, contextual stimuli, and conceptual indicators. The use of contextual texts also contributes to enhancing the instrument’s relevance to students’ real-world experiences. Stimuli related to health technology, renewable energy, environmental issues, and technological applications encourage students not only to recall concepts but also to connect chemical bonding concepts with scientific phenomena in daily life. These findings align with the Organisation for Economic Co-operation and Development’s view that science literacy assessments should emphasize the ability to apply scientific knowledge in meaningful contexts, rather than merely measuring the memorization of concepts.

The use of the four-tier diagnostic format also strengthens the instrument's ability to identify misconceptions more accurately compared to conventional multiple-choice tests. In conventional assessments, students may arrive at the correct answer through guesswork without truly understanding the concept being measured. However, in the four-tier structure, students are required to provide an answer, a scientific rationale, and a confidence level for both. This structure allows for distinguishing between students who truly understand the concept, those who do not understand it, and those who hold misconceptions. Additionally, the inclusion of a confidence level helps reduce false positives situations where students answer correctly but have low conceptual confidence. Similar findings were reported by Treagust (1988), Islami et al. (2019), and Yuberti et al. (2020), who stated that four-tier diagnostic instruments are more effective at distinguishing misconceptions from guessing behavior because they simultaneously evaluate conceptual understanding and confidence in thinking.

The integration of science literacy aspects into the four-tier diagnostic instrument is the primary novelty of this study. Previous research generally focused only on developing four-tier diagnostic instruments to identify misconceptions without explicitly integrating science literacy competencies and contextual applications based on the PISA framework. Unlike previous studies, this research integrates the analysis of misconceptions, contextual scientific phenomena, and science literacy competencies into a single diagnostic assessment framework. The developed instrument not only emphasizes conceptual accuracy but also encourages students to explain scientific phenomena, interpret contextual information, and connect chemical bonding concepts to real-world situations. This integration is crucial because misconceptions in chemistry often arise when students fail to connect abstract concepts at the microscopic level with observable macroscopic phenomena. Therefore, the developed instrument not only serves as a tool for diagnosing misconceptions but also supports the development of students' science literacy skills in alignment with modern science education and PISA-based assessment.

Findings regarding invalid test items also provide insight into the challenges in developing a four-tier diagnostic instrument based on science literacy. Item number 9 was deemed ineffective because its distractor options failed to represent authentic misconceptions regarding VSEPR theory and molecular shapes. Distractors that were too easy to eliminate resulted in the item being less effective at encouraging in-depth conceptual analysis. This indicates that distractors in diagnostic instruments should stem from common, real-world alternative conceptions held by students, rather than merely statements that are clearly incorrect. This finding aligns with David F. Treagust's view that misconception-based distractors are crucial for accurately identifying students' conceptual difficulties.

Similarly, item number 10 is not yet optimal in measuring science literacy competencies because the item emphasizes concept classification over contextual interpretation and scientific reasoning. Although the item used data on the boiling points of hydrogen compounds as a stimulus, the developed question did not fully require students to analyze intermolecular forces and hydrogen bonding phenomena. These findings indicate that integrating science literacy into diagnostic instruments is not sufficient merely by adding contextual text, it also requires alignment between indicators, reasoning processes, and distractor construction with science literacy competencies. Thus, the development of science literacy-based diagnostic instruments is more complex than that of conventional diagnostic instruments because each component must be able to represent conceptual understanding, contextual reasoning, and patterns of misconceptions simultaneously.

Furthermore, item number 22 demonstrates a mismatch between the indicator and the constructed question. The indicator aims to measure students' understanding of the process of ion bond formation through electron transfer, whereas the item emphasizes the ability to recall electron configurations. Consequently, students may answer correctly without truly understanding the ion bond formation process itself. This finding indicates that alignment between indicators, cognitive processes, and item construction is a critical aspect in the development of diagnostic instruments. Without such alignment, test items may fail to measure the competencies they are intended to identify, even if they appear conceptually correct.

Differences in results between CVR and I-CVI analyses were also found in this study. The CVR analysis indicated that some test items were invalid because the values obtained did not exceed the minimum threshold of 0.99 for five validators. Conversely, the I-CVI analysis showed that all items met the minimum threshold of 0.78 according to Polit et al. (2007). This discrepancy arises because CVR places greater emphasis on expert agreement regarding an item's essentiality, whereas I-CVI focuses on an item's relevance based on validator assessments. Consequently, an item may still be considered relevant by most validators but may not meet the stricter essentiality criteria of the CVR. The use of multiple validity approaches in this study provides a more comprehensive evaluation of the developed instrument while helping to identify items requiring revision more accurately.

The results of this study have important implications for chemistry learning and assessment. The developed four-tier science literacy-based diagnostic instrument has the potential to help teachers identify patterns of student misconceptions more comprehensively, particularly regarding abstract chemical bonding concepts. The integration of contextual scientific phenomena can also support contextual learning and help students connect chemical concepts to everyday life. Furthermore, this instrument aligns with current trends in science education that emphasize science literacy competencies and contextual reasoning skills, as reflected in the PISA framework and competency-based curricula.

Nevertheless, this study still has several limitations. The evaluation of the instrument is limited to content validity and has not yet involved large-scale empirical testing, construct validity analysis, or reliability testing on a broader student population. Furthermore, the developed instrument focuses solely on chemical bonding material, so its applicability to other chemistry topics remains unknown. Therefore, future research is recommended to conduct empirical validation, reliability testing, mapping of student misconceptions, and analysis of the instrument's effectiveness in broader learning contexts to strengthen the practicality and generalizability of the developed instrument.

IV. CONCLUSION

This study developed a science-literacy-based four-tier diagnostic instrument for identifying misconceptions in chemical bonding through the planning and development stages. Content validation by five experts provided strong evidence of the instrument's content relevance. Of the 40 items developed, 37 met the CVR criterion of 1 for all assessed aspects and were retained for further testing, whereas three items required substantial revision based on the CVR results. In contrast, all 40 items met the minimum I-CVI criterion of ≥ 0.78. At the instrument level, the overall CVI was 0.9875 and the S-CVI/Ave was 0.99375, indicating a high level of expert agreement regarding the relevance and representativeness of the instrument's content. These findings support the content validity of the developed instrument but should not be interpreted as evidence of its overall psychometric validity or diagnostic accuracy. Further studies should therefore conduct empirical testing with students to examine construct validity, reliability, item functioning, classification accuracy, and the instrument's ability to identify patterns of misconceptions before it is implemented for broader diagnostic use in chemistry education.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest, whether financial, professional, or personal, that could have influenced the research process, interpretation of the findings, or publication of this article.

AUTHOR CONTRIBUTIONS

Junia Marsanita *contributed to the conceptualization, research design, instrument development, data collection, data analysis, and preparation of the original manuscript.* ***Wiwi Siswaningsih*** *contributed to the conceptualization, methodology, supervision, validation, interpretation of the findings, and critical review of the manuscript.* ***Febriati Dian Mubarokah*** *contributed to the methodology, instrument* *validation, data interpretation, and review and revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for the integrity of the work.*

FUNDING
This research received no external funding.

ACKNOWLEDGMENT
The authors would like to express their sincere appreciation to Universitas Pendidikan Indonesia for the academic support provided during this study. The authors also gratefully acknowledge the chemistry education lecturers and high school chemistry teachers who participated as expert validators and provided valuable evaluations and suggestions for the development and content validation of the science-literacy-based four-tier diagnostic instrument.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES
The authors declare that generative artificial intelligence (AI) and AI-assisted technologies were used solely to support language refinement and improve the clarity and readability of the manuscript. These technologies were not used to generate research data, conduct data analysis, interpret the findings, or make scientific conclusions. All AI-assisted content was carefully reviewed, verified, and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

REFERENCES
Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312. https://doi.org/10.1080/09500693.2010.512369

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Arikunto, S. (2016). Prosedur Penelitian Suatu Pendekatan Praktik Penelitian. *In Jurnal Universitas Udayana*. ISSN (Vol. 2302). Rineka Cipta.

Dewi, N. P., Laila, Y., Heffi, R., & Rahmawati, A. (2020). Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Tingkat Tinggi tentang Materi Hereditas untuk Peserta Didik SMA/MA. *JEP (Jurnal Eksakta Pendidikan), 4*(10).

Firman, H. (2013). *Evaluasi Pembelajaran Kimia*. Bandung: UPI Press.

Hendrawani, H. (2023). Katalog Miskonsepsi dalam Pembelajaran Ikatan Kimia. *Empiricism Journal*, *4*(2), 648–656. https://doi.org/10.36312/ej.v4i2.1729

Hofifah, A. H., Aeni, A. N., & Syahid, A. A. (2025). The Influence of TikTok Social Media on the Moral Character (Akhlakul Karimah) of Elementary School Students: Pengaruh Media Sosial Tiktok terhadap Akhlakul Karimah Peserta Didik Sekolah Dasar . *Ainara Journal (Jurnal Penelitian Dan PKM Bidang Ilmu Pendidikan)*, *6*(4), 446–453. https://doi.org/10.54371/ainj.v6i4.1048

Islami, D., Suryaningsih, S., & Bahriah, E. S. (2019). Identifikasi Miskonsepsi Siswa pada Konsep Ikatan Kimia Menggunakan Tes Four-Tier Multiple-Choice (4TMC). *JRPK: Jurnal Riset Pendidikan Kimia*, *9*(1), 21–29. https://doi.org/10.21009/jrpk.091.03

Juliani, R. P., & Erita, S. (2023). Analisis Validitas dan Reliabilitas Instrumen Penilaian Kemampuan Berpikir Kritis dalam Konteks Sekolah Menengah. *JEID: Journal of Educational Integration and Development, 3*(3), 169–179.

Juliartawan, I.G.NK.A. & Riastini, P.N. (2022). Teacher as a Motivator: Gender, Ownership of Educator Certificates, and Age of Elementaru Teachers in Improving Student Motivation during Online Learning. *International Journal of Elementaru Education (IJEE), Vol. 6, No.2,* 182-194.

Kemendikbud Ristek. (2024). *Capaian Pembelajaran Pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, Dan Jenjang Pendidikan Menengah Pada Kurikulum Merdeka. Kementrian Pendidikan, Kebudayaan, Riset dan Teknologi*.

Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology, 28*, 563–575.

Mukhlisa, N. (2023). Validitas Tes. *JUARA SD: Jurnal Pendidikan dan Pembelajaran Sekolah Dasar, 2*, 142–147.

Muliaman, A. ., Sakdiah, H., & Ginting, F. W. . (2022). Analisis Employability Skill dan Literasi Sains Siswa Melalui Authentic Self-Assessment pada Kurikulum Merdeka di SMA Aceh Utara. *JPF (Jurnal Pendidikan Fisika) Universitas Islam Negeri Alauddin Makassar*, *11*(1), 24–32. https://doi.org/10.24252/jpf.v11i1.34010

Nahadi, & Firman, H. (2019). *Asesmen Pembelajaran Kimia*. Bandung: UPI Press.

Nahadi, Siswaningsih, W., Purnawarman, P., Lestari, T., Febriani, A. E., & Rohmawati, T. (2022). Development of Minimum Competency Assessment (AKM) on Chemical Materials. *Moroccan Journal of Chemistry, 10*(3), 452 463.

Noviana, N., Zuhrotul, B. H., & Zakaria, M. . (2025). The Effectiveness of Alphabet Block Games on Language Development in Early Childhood Letter Recognition: Efektivitas Permainan Balok Huruf terhadap Perkembangan Bahasa dalam Pengenalan Huruf Anak Usia Dini. *Ainara Journal (Jurnal Penelitian Dan PKM Bidang Ilmu Pendidikan)*, *6*(3), 348–356. https://doi.org/10.54371/ainj.v6i3.976

OECD. (2023). Pisa 2025 Science Framework. *OECD (Organisation for Economic Co-Operation and Development) Publication*, *May 2023*, 1–93.

Polit, D. F., Beck, C. T., & Owen, S. v. (2007). Focus on research methods: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Pradini, N. L., Rahmad Wijaya, B., & Jannah, A. N. (2022). Analisis Literasi Sains Dalam Upaya Implementasi Pendidikan Abad 21. Eductum: Jurnal Literasi Pendidikan, 1(1), 12–20.

Puspitasari, W. D., & Febrinita, F. (2021). Pengujian Validasi Isi (Content Validity) Angket Persepsi Mahasiswa terhadap Pembelajaran Daring. *Focus Action Of Research Mathematic, 4*(1), 77–90.

Rahayu, D. S., & Fitriza, Z. (2021). Identifikasi Miskonsepsi Peserta Didik Pada Materi Ikatan Kimia : Sebuah Studi Literatur. *EDUKATIF : JURNAL ILMU PENDIDIKAN*, *3*(3), 1084–1091. https://doi.org/10.31004/edukatif.v3i3.510

Ramadhan, M. F., Siroj, R. A., & Afgani, M. W. (2024). Validitas and Reliabilitas. *Journal on Education 06*(02), 10967–10975.

Rohmaya, N. (2022). Peningkatan literasi sains siswa melalui pembelajaran IPA berbasis socioscientific issues (SSI). *Jurnal Pendidikan MIPA, 12*(2), 107–117. https://doi.org/10.37630/jpm.v12i2.553

Sensus, M., Arifin, K., & Munir, A. (2022). Validitas Soal pada Asesmen Kompetensi Minimum Materi Ekologi SMA Kelas X. *Wahana-Bio: Jurnal Biologi dan Pembelajarannya, 14*, 1–10.

Septikasari, R., & Frasandy, R. N. (2018). Keterampilan 4C Abad 21 dalam Pembelajaran Pendidikan Dasar. *Jurnal Tarbiyah Al-Awlad Vol. (8) Edisi 2*, 122.

Sudiatmika, A. A. I. A. R., Subagia, W., & Utari, P. H. (2025). PENGEMBANGAN INSTRUMEN TES DIAGNOSTIK FOUR TIER UNTUK MENGIDENTIFIKASI MISKONSEPSI PESERTA DIDIK. In *Jurnal Pendidikan dan Pembelajaran IPA Indonesia* (Vol. 15, Issue 1).

Suryadi, T., Alfiya, F., Yusuf, M., Indah, R., Hidayat, T., & Kulsum, K. (2023). Content Validity for the Research Instrument Regarding Teaching Methods of the Basic Principles of Bioethics. *JPKI: Original Research, 12*(2), 186–202.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconception in science. *International Journal of Science Education, 10*, 159-169. https://doi.org/10.1080/0950069880100204

Widianti, Y., & Pratikno, A. S. (2024). Analisis Penggunaan Media Baca Literacy Cloud terhadap Minat Baca Peserta Didik Sekolah Dasar. *Ainara Journal (Jurnal Penelitian Dan PKM Bidang Ilmu Pendidikan)*, *5*(3), 247–254. https://doi.org/10.54371/ainj.v5i3.478

Yasthophi, A., Ritonga, S. P. (2017). Miskonsepsi Mahasiswa Mengenai Ikatan Ion dalam Senyawa NaCl. *Jurnal Tadris Kimiya* (Vol. 1, Issue 2).

Yuberti, Y., Suryani, Y., & Kurniawati, I. (2020). Four-Tier Diagnostic Test with Certainty of Response Index to Identify Misconception in Physics. *Indonesian Journal of Science and Mathematics Education*, *3*(2), 245–253. https://doi.org/10.24042/ijsme.v3i2.6061

Yuliati, Y. (2017). LITERASI SAINS DALAM PEMBELAJARAN IPA. *Jurnal Cakrawala Pendas*, *3*(2). https://doi.org/10.31949/jcp.v3i2.592

Yusup, F. (2018). UJI Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23.