Jurnal Pendidikan MIPA
Volume 16. Nomor 4, Tahun 2026 | ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Development and Validation of a Reading and Numeracy Literacy Assessment Instrument on Stoichiometry Using LMS

Anugrance Citra Rahma Setiyani[1)], Triannisa Rahmawati[1),*], Nahadi[1)], Wiwi Siswaningsih[1)]

[1)]Chemistry Education, Indonesia University of Education

*Coresponding Author: triannisa.ra@upi.edu

ABSTRACT

According to PISA results, Indonesian students' reading and numeracy skills remain low. Therefore, a quality assessment tool is needed to evaluate students' reading and numeracy skills, especially in stoichiometry. The purpose of this study was to determine an appropriate instrument for assessment. The research method used was test design and development, as outlined in the Standards for Educational and Psychological Testing. The results showed that the instrument had adequate characteristics in content validation, empirical validity, reliability, difficulty level, and discriminatory power, with most falling into the good category. Therefore, this instrument can be used as an assessment to obtain evidence of the instrument's validity and psychometric characteristics. Further research is recommended to evaluate students' levels of critical thinking, reading literacy, and numeracy. In addition, the instrument can be tested on a more diverse sample and in more schools, expanding it to other chemistry materials.

Keywords: Assessment Instruments; Reading Literacy; Numeracy Literacy; and Stoichiometry

INTRODUCTION

To address the challenges of life and the workplace, as well as the more sophisticated advancements in science and technology, 21st-century education requires students to possess a variety of competencies. The Partnership for 21st Century Skills (P21) identifies communication, collaboration, critical thinking, and creativity as essential competencies that students must possess in the 21st century (Partnership for 21st Century Skills, 2019). To support mastery of these skills, students need to have adequate literacy, as these skills are the basis for understanding information, analyzing problems, and making logical, critical decisions (Pratama et al., 2025).

In the framework of 21st century education, there are two types of literacy, namely reading literacy and numeracy literacy (S. R. Utami & Kusmayati, 2024). The need to improve these two abilities is highlighted by the Programme for International Student Assessment (PISA) (2022) findings, which show that Indonesian students' literacy achievement remains low and tends to fluctuate from year to year. According to the 2022 PISA results, Indonesian students continue to perform below the OECD average in reading, math, and science. In mathematics, only 18% of students met the required competency level, while in science and reading, only 25% and 34% did so, respectively (OECD, 2022). These poor results highlight the necessity of ongoing efforts to improve reading and numeracy, as these two skills are crucial for learning, for helping students develop higher-order thinking abilities, and for preparing them for the challenges of the twenty-first century (Amelia et al., 2024). Based on these findings, it can be seen that students' basic abilities in understanding the information they receive and solving problems that occur in everyday life, especially in the field of numeracy or mathematics, are still low (Kamaludin, 2024). These results highlight the need for continued efforts to improve students' reading and numeracy skills, including the provision of assessment tools that reliably and accurately measure competency. Therefore, the Indonesian government has established a new policy known as the Minimum Competency Assessment (AKM) to replace the National Examination (UN) (Nahadi et al., 2022).

Minimum Competency Assessment (AKM) is a basic competency assessment required of all Indonesian students to help them develop their abilities and play an active role (Rohim et al., 2021). This assessment is used to measure the results of students' cognitive learning in their literacy skills, especially reading literacy and

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

numeracy (Novita et al., 2021). Literacy is a fundamental skill that is valued, especially for understanding the various information encountered in everyday life and technological developments (Rohman et al., 2022). Reading literacy is a fundamental ability that enables students to comprehend, analyze, and extract information from texts to enhance their knowledge, critical thinking, and language proficiency, and to support learning success across a variety of professions (Ambarita et al., 2021; Cicilia & Nursalim, 2019). Meanwhile, numeracy literacy refers to a person's capacity to use mathematical knowledge to evaluate data, solve various contextual problems, and use analytical findings to inform decision-making (Amidi, 2024). Therefore, it is important to develop students' critical thinking skills to be useful in solving problems and making decisions (Geiger & Schmid, 2024).

One approach to improving the quality of reading and numeracy is the contextual approach (Hidayat et al., 2024). A contextual approach is a learning process that links subject matter to real-life contexts to increase student motivation and serves as the foundation for an efficient and effective teaching and learning process (N. D. Utami, 2024). This approach aligns with findings Brinus et al. (2019), which suggest that a contextual approach to learning is effective in improving students' understanding of the material. Furthermore, Parhan & Sutedja (2019) suggest that this contextual approach allows students to develop skills in connecting the material to everyday life.

Chemistry education plays a crucial role in integrating environmental, social, and economic aspects into development strategies (Hernani et al., 2024). Stoichiometry is one of the chemistry topics chosen for this study because it is highly relevant to real-life applications. This context and approach align with the principles of the current independent curriculum, which strongly emphasizes a deep learning approach, namely meaningful, in-depth, and contextual learning to promote 21st-century learning (Handayani et al., 2025; Wirati, 2023).

Various studies have developed assessment instruments based on AKM and scientific literacy in chemistry learning. Jayanti et al. (2026) developed an AKM instrument for chemistry based on the Minimum Competency Assessment framework, while Adisti Yuliastrin et al. (2025) developed and validated a scientific literacy instrument for secondary school students. However, these studies still focus on the development of AKM instruments or scientific literacy in general and have not specifically integrated reading literacy and numeracy literacy into a single assessment instrument for stoichiometry. Furthermore, the implementation of instruments through a Learning Management System (LMS) as a digital assessment medium and analysis of instrument quality, including content validity, empirical validity, reliability, difficulty level, and discriminatory power, have also not been comprehensively reported. On the other hand, several studies indicate that stoichiometry remains one of the most difficult chemistry topics to understand because it requires mastery of prerequisite concepts as well as good numeracy skills. Students' low understanding of this material indicates the need for assessment instruments capable of more accurately identifying reading literacy and numeracy skills to support the learning evaluation process.

Based on this gap, this research offers a novel approach through the development of an assessment instrument that integrates reading literacy and numeracy in stoichiometry based on the AKM framework. The instrument was implemented using a Moodle-based LMS, allowing for more systematic, flexible, and efficient assessment. Furthermore, the instrument's quality was comprehensively evaluated through analyses of content validity, empirical validity, reliability, difficulty level, and discriminatory power. Thus, this research not only produces a validated digital assessment instrument but also provides an evaluation alternative that aligns with the characteristics of stoichiometry material and the demands of 21st-century learning.

To ensure that the instruments developed are truly suitable for use as assessment tools, the quality of the instruments needs to be thoroughly evaluated. The quality of an assessment instrument is a crucial aspect because it generates data and variables (Salim & Juita, 2025). Analyzing the quality of a test instrument is crucial for creating a reliable and valid instrument to improve the quality of learning (Setiyawan & Wijayanti, 2020). Instruments that do not meet quality standards for measurement instruments may produce inaccurate data, which could affect student learning outcomes. Therefore, an analysis of the instrument's characteristics, including validity, reliability, difficulty level, and discriminative power, is necessary to ensure that the instrument can be used as a quality measurement tool.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

METHOD

According to the Standards for Educational and Psychological Testing, the research strategy employed in this study is a test design and development approach (AERA et al., 2014). In this study, this method is applied to the development and validation stages of Moodle-based Learning Management System (LMS) reading literacy and numeracy assessment instruments, which are then tested for validity, reliability, level of Difficulty, and discriminatory power. There are four phases of this research design: 1) the development and evaluation stage of test specifications; 2) the development, trial, and evaluation stage of test items; 3) the practice stage and evaluation of new test forms; 4) the development stage of procedures and materials for administration and assessment. The participants were 35 students of grades X-5 and 31 students of grade X-7 at one of SMA 1 Subang, specifically those who had studied stoichiometry material. In addition, this study involved five validators: three lecturers from the Chemistry Education study program at the Faculty of Mathematics and Natural Sciences, University of Indonesia (FPMIPA UPI) and two high school chemistry teachers. The research procedure is presented in Figure 1.

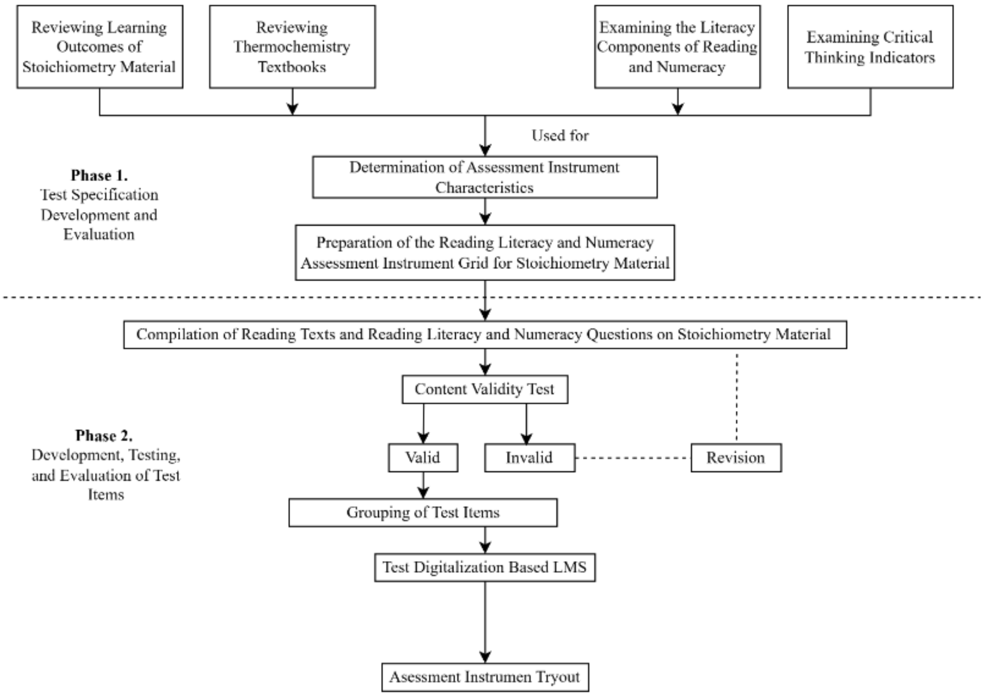


Figure 1. Research Procedure

The assessment of stoichiometry learning outcomes, the content of stoichiometry material from various textbook sources, and a review of the literature on minimum competency assessments and critical thinking indicators served as the foundation for developing this instrument. A draft assessment tool was developed based on this flow. To ensure that learning objectives and the elements to be tested are aligned, this draft serves as the basis for generating test items. Initial specifications, such as item format, time allocation, target participants, and test administration procedures, are determined.

The grid was developed before drafting the instrument by integrating reading and numeracy literacy indicators, based on the AKM framework, into the stoichiometry competency. Each literacy indicator was translated into question indicators tailored to the material's characteristics, so that each item not only measured mastery of chemical concepts but also the ability to understand information, interpret data, and apply quantitative reasoning in solving stoichiometry problems. The draft instrument was then created using reading and numeracy literacy questions related to stoichiometry.

This research focused on the development and validation of assessment instruments. Therefore, data analysis in this study only aimed to evaluate the psychometric characteristics of the developed instruments, including content validity, empirical validity, reliability, difficulty level, and discriminatory power based on student responses. Analysis of students' reading literacy, numeracy literacy, and critical thinking skills was not conducted as they were beyond the scope of this study.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Following the series' creation, content validation was conducted through expert evaluation to assess each item's suitability, relevance, and clarity with respect to its indicators. The validity of the instrument content was analyzed using the Content Validity Ratio (CVR) method (Selaras et al., 2019). The coefficient obtained is then matched with the minimum limit of the CVR value proposed by Lawshe (1975).

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \tag{1}$$

Furthermore, empirical validity is a type of validity that uses other measuring instruments as a comparison or benchmark to assess the extent to which an instrument is valid (Arikunto, 2016). The product-moment correlation equation was used to assess the items' empirical validity (Karira et al., 2023). The resulting correlation coefficient is interpreted based on the criteria proposed by Sari et al. (2023).

$$r_{xy} = \frac{N.\sum XY - (\sum X).(\sum Y)}{\sqrt{(N.\sum X^2 - (\sum X)^2).(N.\sum Y^2 - (\sum Y)^2)}} \tag{2}$$

Instrument reliability was analyzed using the KR-20 coefficient for dichotomously scored items and the Cronbach's Alpha coefficient for polytomously scored items. An instrument is considered reliable if it has a minimum reliability coefficient of 0.70 (Fraenkel et al., 2012; Gliem & Gliem, 2003).

$$\rho_{KR20} = \frac{k}{k-1}\left[1 - \frac{\sum pq}{\sigma^2}\right] \tag{3}$$

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_t^2}\right) \tag{4}$$

Difficulty level analysis is carried out to determine the difficulty category for each question item, including easy, medium, and difficult (Fatimah and Alfath, 2019a). Then, the results of the calculation of the level of difficulty are interpreted based on the criteria put forward by Asbupel et al. (2021).

$$P = \frac{B}{JS} \tag{5}$$

A power analysis was conducted to assess the ability of each question item to differentiate between high- and low-ability students. Students were grouped using the upper- and lower-group methods, each comprising 50% of the total number of students, because the sample size was less than 100. The results of the discriminatory power analysis were then classified into the categories very good, good, adequate, poor, and very poor (Fatimah and Alfath, 2019a). And the results of the discriminatory power are interpreted based on the categories proposed by Erfan et al. (2020).

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B \tag{6}$$

This study still uses the Classical Test Theory (CTT) approach to evaluate instrument quality. Therefore, further research is recommended to use modern approaches, such as the Rasch Model or Item Response Theory (IRT), so that instrument characteristics can be evaluated more comprehensively.

RESULTS AND DISCUSSION

In the first phase, namely the development and evaluation of test specifications, a study was conducted on the learning outcomes of senior high schools for stoichiometry material based on the decree of the head of the education standards, curriculum, and assessment agency of the Ministry of Primary and Secondary Education number 46/H/KR/2025 concerning learning outcomes in early childhood education, primary education levels, and secondary education levels. The purpose of this analysis was to identify the primary materials to be used in the research, while also matching the depth and breadth of the materials to the design of the assessment instrument focused on measuring reading literacy and numeracy. Based on the study's results, the learning outcomes and scope of the stoichiometry material have been identified, as presented in Table 1.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Tabel 1. Learning outcomes and scope of the stoichiometry material

Learning Outcome	Scope of Stoichiometry
Students have the ability to apply stoichiometric concepts in various quantitative aspects of chemical reactions.	1. Calculation of relative atomic mass (Ar) and relative molecular mass (Mr) 2. Concept of mole 3. Empirical and molecular formulas 4. Calculation of the content of an element in a compound 5. Calculation in chemical equations 6. Limiting reagent 7. Determination of the chemical formula of a hydrate compound 8. Solution concentration units

The depth of the material is related to the scope of the knowledge dimensions that students must learn and is systematically arranged based on the relationships among concepts to make it easy to understand (Wulan, 2023). Thus, the alignment between assessment items and curriculum objectives needs to be considered, as a mismatch can hinder learning, reduce student motivation, and reveal a gap between the assessment instruments used and the competencies targeted in the curriculum (Khan et al., 2025).

The scope of stoichiometry was comprehensively reviewed by referring to high school chemistry textbooks and general chemistry textbooks to ensure that the developed assessment instruments aligned with the applied curriculum and maintained conceptual accuracy. This step was crucial to ensure the chemistry concepts in the questions aligned with those in the chemistry textbooks to avoid misunderstandings (Metafisika, 2018). Mastery of concepts was necessary for students to fully understand chemistry concepts and use them as a foundation for learning more complex concepts (Yustiqvar et al., 2019). Furthermore, this stage served as the foundation for the instrument development process, ensuring that each question item represented the indicators, material, and competencies being measured, and that the instrument had good content feasibility (Sumarno, 2020).

The development of reading and numeracy literacy indicators is based on the AKM level 5 framework aimed at grade X students. The content used consists of informative texts in everyday life. In reading literacy, the competencies measured include: (1) access and retrieve ;(2) interpret and integrate; and (3) evaluate and reflect. Meanwhile, in numeracy literacy, the data and uncertainty domains are applied, specifically the data and representation subdomains. The cognitive levels of numeracy literacy are as follows: (1) knowing; (2) applying; and (3) reasoning. Therefore, the development of this instrument is expected to assess students' abilities in reading and numeracy literacy according to established criteria. The critical thinking ability indicators proposed by Facione (2015) are interpretation, analysis, evaluation, inference, explanation, and self-regulation (Arini & Rahayu, 2023). However, in this study, four critical thinking indicators were used, and their descriptions are presented in Table 2.

Tabel 2. Critical thinking indicator description

Indicator	Critical Thinking Indicator Description
Interpretation	The interpretation indicator explains how someone interprets information from their point of view. In this indicator, students can record their thoughts and observations.
Analysis	The analysis indicator describes the process of observing an object to obtain relevant facts. Following their presentation, students will carry out an inquiry or observation and compile responses that they believe are pertinent or correct (in line with the question or reasoning).
Evaluation	The evaluation indicator is the activity of reinterpreting the answers from the meaning analysis stage to identify which are relevant or accurate
Inference	This stage's conclusion indication is still connected to the one before it, where students will make inferences based on the answers they have selected

As a result, it is anticipated that this test will assess students' critical thinking abilities in addition to their reading and math proficiency.

The results of the study were used to develop a test framework that serves as a systematic guideline for creating test items. The framework was designed based on the Minimum Competency Assessment (AKM) framework and critical thinking skill indicators, so that each test item can represent the competencies being measured: reading literacy, numeracy literacy, and critical thinking skills. The test framework design includes

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

several important components, such as text titles, subtopics, question indicators, indicators of reading literacy and numeracy, indicators of critical thinking skills, question format, and item numbers. In addition, the test specification stage also includes determining instrument characteristics, test format, time allocation, target participants, and test implementation procedures. The instrument characteristics examined in this study include content validity, empirical validity, reliability, level of difficulty, and discriminatory power. These five aspects were analyzed to ensure that the developed instrument can measure the intended competencies accurately and consistently and has adequate item quality for use in the assessment process (Saputri et al., 2023).

This study analyzed the instrument's quality, including content validity, empirical validity, reliability, difficulty level, and discriminatory power. However, this study did not examine the level of achievement in reading literacy, numeracy literacy, or critical thinking skills, as measured by the instrument. Therefore, further research is needed to evaluate the instrument's ability to measure these three aspects, thereby enabling a more comprehensive review of its quality. The test format was designed according to the characteristics of the test items used, namely, a presentation of information text followed by a series of questions. The time allocation for completing the test was determined based on the number and level of complexity of the test items developed. The target participants in this study were X-grade high school students who had studied stoichiometry. The test was administered after the learning process was completed, allowing time to complete it within the predetermined duration. Then, the results of the trial were used to analyze the instrument's empirical validity, reliability, difficulty level, and discriminatory power.

The second phase of this research involved the systematic development of reading and numeracy literacy items based on the established test specifications. This process involves designing reading texts related to stoichiometry that serve as stimuli or prompts for students to answer the questions presented. Examples of the developed informative texts are shown in Figures 2 and 3.

CAIRAN ANTISEPTIK

Gambar 4. Penggunaan Antiseptik

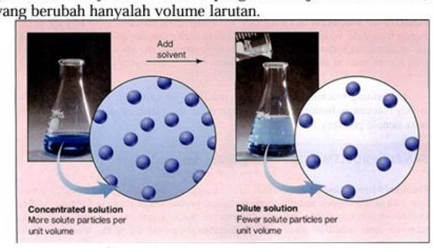
Sumber : https://www.alodokter.com/lebih-jauh-tentang-efektivitas-antiseptik

Antiseptik merupakan cairan berbahan kimia yang digunakan untuk menghambat pertumbuhan dan membunuh mikroorganisme, seperti bakteri, virus, dan jamur, sehingga dapat mencegah terjadinya infeksi. Antiseptik banyak dimanfaatkan untuk membersihkan luka ringan, mensterilkan kulit sebelum antiseptik medis, serta menjaga kebersihan tangan dan permukaan kulit. Dalam praktik medis, antiseptik tersedia dalam berbagai bentuk, seperti cairan, gel, semprot, dan krim.

Sebagian besar cairan antiseptik yang beredar menggunakan bahan aktif berbasis antiseptik, terutama alkohol (massa molar 46 gram/mol) dengan kadar tertentu, misalnya 95% atau 70%. Selain antiseptik, beberapa senyawa lain yang juga digunakan sebagai antiseptik antara lain antiseptik peroksida, chlorhexidine, chloroxylenol, dan iodofor. Senyawa-senyawa tersebut bekerja dengan merusak struktur sel mikroorganisme sehingga menghambat pertumbuhannya.

Efektivitas antiseptik tergantung pada konsentrasinya. Jika terlalu encer, antiseptic tidak akan membunuh kuman. Jika terlalu pekat, bisa merusak kulit. Karena itu, pembuatan antiseptik harus melalui perhitungan yang tepat.

Dalam ilmu kimia, larutan itu terbentuk ketika zat terlarut (misalnya, alkohol) bercampur merata ke dalam pelarut (misalnya, air). Konsentrasi menyatakan seberapa banyak zat terlarut dalam sejumlah larutan. Yang menarik, konsentrasi tidak bergantung pada volume total larutan. Artinya, satu tetes dan satu ember larutan yang sama memiliki konsentrasi yang sama.
Larutan pekat (dengan kadar alkohol tinggi) dapat diencerkan menjadi lebih encer dengan menambahkan pelarut, misalnya air. Pada saat pengenceran, jumlah zat aktif (dalam satuan mol) tidak berubah, yang berubah hanyalah volume larutan.



Gambar 5. Larutan Pekat dan Larutan Encer
Sumber: Silberberg, 2010.

Figure 2. Informational text *cairan antiseptik*

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Bayangkan dua larutan: larutan pekat dan larutan encer. Dalam larutan pekat, partikel zat terlarut (misalnya molekul alkohol) sangat padat per satuan volume. Ketika kita menambahkan pelarut (air), volume larutan bertambah, tetapi jumlah partikel zat terlarut tetap sama. Akibatnya, dalam volume yang sama, larutan encer mengandung lebih sedikit partikel zat terlarut dibandingkan larutan pekat.

Karena itu, di laboratorium atau industri, sering disiapkan larutan stok, yaitu larutan yang sengaja dibuat pekat, lalu disimpan dan diencerkan sesuai kebutuhan. Contohnya, antiseptik alkohol 70% diperoleh dari pengenceran larutan stok alkohol 95% dengan menambahkan pelarut (air). Maka, kadar 70% pada alkohol berarti 70 mL alkohol murni dalam setiap 100 mL larutan (persen volume per volume atau %v/v).

Selain konsentrasi, penyimpanan antiseptik juga memengaruhi kualitas dan keamanannya. Penyimpanan yang tidak tepat dapat menyebabkan kontaminasi mikroorganisme atau perubahan konsentrasi zat aktif akibat penguapan, terutama pada antiseptik berbahan dasar antiseptik. Oleh karena itu, antiseptik sering dikemas dalam wadah tertutup rapat atau kemasan sekali pakai untuk menjaga kemurnian dan efektivitasnya.

Sumber :
https://www.alodokter.com/lebih-jauh-tentang-efektivitas-antiseptik
https://www.alodokter.com/6-cairan-antiseptik-yang-ampuh-bunuh-kuman
Silberberg, M. (2010). *Principles of General Chemistry Second Edition*. New York: McGraw-Hill.

Figure 3. Informational text *cairan antiseptik*

Following the compilation of the information text, 50 reading literacy and numeracy test items were developed, referring to the previously prepared test grid and the Minimum Competency Assessment (AKM) framework developed by the Assessment and Learning Center. To ensure congruence among the targeted literacy features, the question indicators, and the competencies being measured, item development was conducted in a methodical manner. Five different forms were used to present the items: matching, short response, multiple-choice, complex multiple-choice, and descriptive. The interaction among the question indicators, the reading and numeracy literacy components, and the critical thinking skill indicators served as the basis for developing each item. In order to ensure that the instrument could more accurately reflect the range of student skills, the item creation also considered the level of cognitive processes being measured. Thus, each item not only serves to measure reading literacy and numeracy skills but also accommodates the achievement of the established critical thinking skill indicators. Before being used in the trial phase, all items underwent a content validation process to assess the alignment between the stimulus text, the question indicators, and the critical thinking skill indicators being measured. The results of the instrument content validation are presented in Table 3.

Tabel 3. Overall CVR Value

Content Validity Category	Number of valid test items based on CVR	Description
Text Conformity to Test Items	49	49 test items valid with revisions and 1 test item invalid
Conformity of Test Item Indicators to Test Items	41	41 test items valid with revisions and 9 test items invalid
Conformity of Critical Thinking Indicators to Test Items	42	42 test items valid with revisions and 8 test items invalid

The results in Table 3 indicate that the established constructs can be measured by most test items. However, some items remain irrelevant. The clarity of the test wording, the complexity of the stimulus, and differences in students' ability to comprehend the information presented can all influence this. Therefore, invalid items should be revised before broader implementation.

Based on the results, the test items that met the criteria were then revised according to the validator's suggestions and arranged into two instrument packages, namely Package A and Package B. The two packages were prepared by considering the equivalence of the indicators of reading literacy, numeracy literacy, and critical thinking skills measured, ensuring that both packages have equivalent capabilities. Each package consists of 25 test items, including some of the same items and some different ones, but still measures equivalent indicators. The instrument was packaged in two packages to obtain a variety of test items without reducing the representation of the competencies being measured. The distribution of test items based on the indicators of reading literacy, numeracy literacy, and critical thinking skills in the two packages is presented in Table 4.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Tabel 4. The distribution of items for each critical thinking skill indicator, reading literacy, and numeracy skills

Critical Thinking Indicator	Reading Literacy	Numeracy Skills	Total
Interpretation	2	10	12
Analysis	2	2	4
Evaluation	2	3	5
Inference	2	1	3
Total	8	16	25

Thus, the resulting instrument consisted of 25 questions, including 5 designed information texts, divided into 8 items for reading literacy and 16 for numeracy literacy. Of these items, 12 measured interpretation indicators, 3 measured analysis, 3 measured inference, and 5 measured evaluation as part of critical thinking skills. The development of the test items was divided into 5 question formats: multiple choice, complex multiple choice, matching, short answer, and essay. In addition to item development, an assessment rubric was also developed to ensure the scoring process was carried out objectively, systematically, and consistently.

Table 5. Example the Draf Instrument

<table>
<tr><th colspan="2" rowspan="2">Draft Instrument</th><th colspan="2">Number</th></tr>
<tr><th>A</th><th>B</th></tr>
<tr><td colspan="2">Teks 2 – Metana: Komponen Utama Gas Alam</td><td></td><td></td></tr>
<tr><td>Indikator Butir Soal</td><td>Menilai kebenaran pernyataan mengenai massa, volume, jumlah molekul, dan jumlah mol air yang dihasilkan dari pembakaran sempurna metana pada STP</td><td rowspan="3">6</td><td rowspan="3">6</td></tr>
<tr><td>Indikator Berpikir Kritis</td><td>Analisis</td></tr>
<tr><td colspan="2">Literasi Numerasi
Berdasarkan teks "Metana: Komponen Utama Gas Alam", jika sebanyak 8 gram gas metana (CH_4) dibakar sempurna pada kondisi STP, maka ...
□ Massa oksigen yang diperlukan untuk pembakaran sempurna adalah 32 gram.
□ Massa karbon dioksida yang dihasilkan adalah 22 gram.
□ Volume karbon dioksida yang dihasilkan pada STP adalah 11,2 liter.
□ Jumlah molekul air yang dihasilkan adalah $6{,}02 \times 10^{23}$ molekul.
□ Jumlah mol air yang dihasilkan sama dengan jumlah mol metana yang dibakar.</td></tr>
<tr><td>Indikator Butir Soal</td><td>Menginterpretasikan jumlah mol gas metana (CH_4) pada keadaan STP</td><td rowspan="3">7</td><td rowspan="3"></td></tr>
<tr><td>Indikator Berpikir Kritis</td><td>Interpretasi</td></tr>
<tr><td colspan="2">Literasi Numerasi
Berdasarkan teks "Metana: Komponen Utama Gas Alam", pada keadaan STP 44,8 gram gas metana (CH_4) setara dengan .... mol.</td></tr>
<tr><td>Indikator Butir Soal</td><td></td><td rowspan="3"></td><td rowspan="3">7</td></tr>
<tr><td>Indikator Berpikir Kritis</td><td>Interpretasi</td></tr>
<tr><td colspan="2">Literasi Numerasi
Berdasarkan informasi tabel data atom relatif pada teks "Metana: Komponen Utama Gas Alam", jika 7,2 gram CO_2 dihasilkan dari pembakaran metana dan diukur pada STP. Jumlah mol CO_2 tersebut adalah ... mol</td></tr>
<tr><td>Indikator Butir Soal</td><td>Menilai kebenaran pernyataan mengenai pembakaran metana, hubungan massa metana dengan massa CO_2, peningkatan konsentrasi metana, dan kontribusi gas alam terhadap emisi CO_2.</td><td rowspan="3">9</td><td rowspan="3">9</td></tr>
<tr><td>Indikator Berpikir Kritis</td><td>Analisis</td></tr>
<tr><td colspan="2">Literasi Membaca</td></tr>
</table>

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Berdasarkan reaksi pembakaran sempurna metana (CH_4) dan data peningkatan konsentrasi metana di atmosfer serta emisi karbon dioksida global pada teks 2, pilihlah pernyataan yang benar dibawah ini!

Pernyataan		Jawaban
a)	Pembakaran sempurna metana menghasilkan karbon dioksida dan uap air sesuai dengan perbandingan mol pada persamaan reaksi.	Benar/Salah
b)	Semakin besar jumlah metana yang dibakar, semakin besar pula massa karbon dioksida yang dihasilkan.	Benar/Salah
c)	Peningkatan konsentrasi metana di atmosfer selalu menyebabkan penurunan emisi karbon dioksida global.	Benar/Salah
d)	Massa karbon dioksida hasil pembakaran metana lebih besar daripada massa metana yang dibakar.	Benar/Salah
e)	Pembakaran metana tidak berkontribusi terhadap emisi gas rumah kaca karena metana habis bereaksi.	Benar/Salah
f)	Gas alam dapat berkontribusi terhadap emisi CO_2 baik melalui pembakaran maupun kebocoran metana.	Benar/Salah

After the packaging process was completed, the instruments were implemented in a Moodle-based Learning Management System (LMS) as the assessment medium. Moodle was chosen because it offers various features that support integrated learning and assessment, such as material presentation, question delivery, time management, automatic response collection, and assessment result storage in a single platform. In this study, students first accessed the stoichiometry material available on the LMS, then completed the assessment instruments according to the predetermined package. All student responses were automatically recorded by the system, simplifying the documentation and analysis of assessment results.

Using a Moodle-based LMS offers several advantages over paper-based assessments. In addition to allowing students to access the instruments online anytime and anywhere, Moodle also simplifies question distribution, reduces paper use, and increases assessment efficiency through automatic time recording and response collection. This makes the assessment process more systematic, flexible, and efficient. These findings align with Dongoran and Hariani (2025); Sudiana (2016), who stated that LMS-based learning environments facilitate learning and evaluation because they can be accessed according to student needs.

Furthermore, Moodle not only functions as a medium for delivering questions but also provides various features that support the learning process, such as material presentation, discussion forums, quizzes, assignments, and the integration of various learning media, including text, images, audio, and video. These various features enable students to learn independently or with guidance according to their learning needs. Budhayanti (2023) also reported that utilizing Moodle's features can enhance student learning activities because they can access materials, participate in discussions, and take quizzes all within a single, integrated platform.

This research finding is also supported by Risti (2025), who stated that Moodle is an e-learning platform with advantages such as flexibility, accessibility, data security, and the ability to adapt to user needs. These advantages make Moodle an effective LMS for supporting modern learning and the implementation of digital-based assessments.

In the context of chemistry learning, particularly in stoichiometry, Moodle serves not only as a medium for assessment but also as a learning environment that integrates learning materials with evaluation activities in a single platform. This integration allows students to study the material before working on the assessment instruments, thus ensuring a continuous learning and evaluation process. These results align with research by Janah and Kurniawan (2025), which demonstrated that a Moodle-based LMS can create more varied and engaging learning experiences, especially for students accustomed to using digital devices in their learning process.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Figure 4. Moodle LMS Main Page View

Figure 4 shows the instrument display in the Moodle LMS, consisting of the main page, work instructions, stoichiometry material, and assessment pages. Students first study the available material, then select a question package according to their assigned class, work on the questions within the allotted time limit, and submit their answers through the system. All student work is automatically saved in the LMS, facilitating data recapitulation and analysis.

Therefore, instrument trials are necessary to demonstrate that each test item possesses appropriate characteristics before use as an assessment tool. This aligns with research on the development of scientific literacy instruments, which states that a trial process is necessary to evaluate the validity, reliability, difficulty level, and discriminatory power of each test item, ensuring the instrument's quality before use in learning (Adisti Yuliastrin et al., 2025). Empirical validity is a crucial indicator to ensure that each test item truly measures the intended competency based on student responses. This method is employed because empirical validity assesses how test results relate to criteria relevant to the construct being measured but not to the test itself (Arifin, 2012). The overall empirical validity percentage results are presented in Table 6.

Tabel 6. Percentage of Empirical Validity of Package A and Package B

Category	Number of Questions Items in Package A and Percentage (%)	Number of Questions Items in Package B and Percentage (%)
High	6 (24%)	7 (28%)
Fair	14 (56%)	12 (48%)
Low	5 (20%)	6 (24%)
Total Questions	25	

Based on Table 5, the majority of items in Package A fall into the "sufficient" category (56%), while those in Package B are dominated by the "sufficient" category (48%). Furthermore, 24% of items in Package A and 28% in Package B fall into the "high" category. These results indicate that most items have a fairly strong to strong relationship with the total score, thus effectively measuring the established constructs, namely reading literacy and numeracy in stoichiometry.

However, several items still have low empirical validity, namely items 3, 11, and 21 in Package A and items 6 and 16 in Package B. These items indicate that the items' ability to measure the established constructs is not optimal and require revision before reuse. This condition can be influenced by several factors, such as unclear question wording, inappropriate difficulty levels, the effectiveness of distractors, and the characteristics of stoichiometry material that demands a high level of conceptual understanding and reasoning skills. Research by Rahmawan and Ashfarini (2024), also showed that stoichiometry remains a topic prone to misconceptions among students. This can influence students' response patterns to test items, resulting in low empirical validity for some items.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

The results of this study also indicate that validity obtained through expert assessment does not always guarantee good empirical validity for a test item. In other words, test items that are theoretically stated to align with the indicators and constructs being measured still need to be tested on students to ensure they truly measure the expected abilities. This finding aligns with Karira et al. (2023), who stated that validity based on expert assessment needs to be supported by empirical testing so that the instrument's quality can be evaluated more comprehensively. Therefore, validity testing not only serves to select the appropriate test items for use but also serves as a basis for ensuring that the instrument produces accurate information regarding the abilities being measured. This finding aligns with research by Puspitasari et al. (2025) which states that the instrument validation process is a crucial step in ensuring that the instrument accurately measures the intended competencies before being used in assessment activities.

In the context of chemistry learning, particularly in stoichiometry, empirical validity testing is a crucial step because it can identify test items that are not functioning optimally. This information can be used as a basis for improving the wording, context, and complexity of the questions so that the resulting instrument is truly capable of accurately measuring students' reading and numeracy skills.

After validity testing, the next step in assessing instrument quality is reliability testing. The extent to which a test produces consistent results when administered to the same group is called reliability (Sugiono et al., 2020). Figure 5 for Package A and Figure 6 for Package B show bar charts showing reliability values.

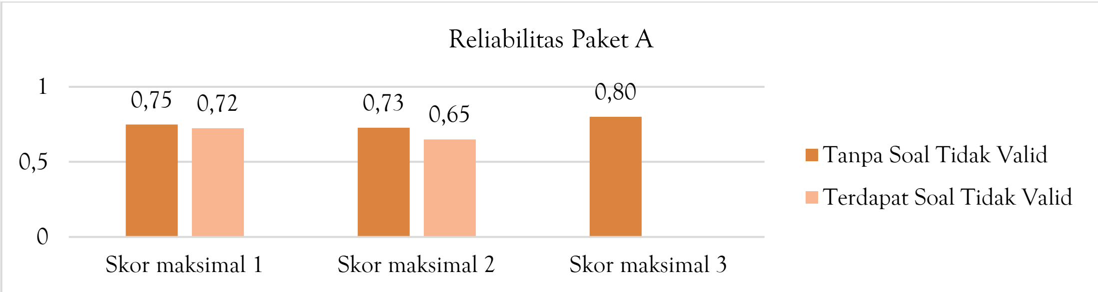


Figure 5. Package A Reliability Bar Chart

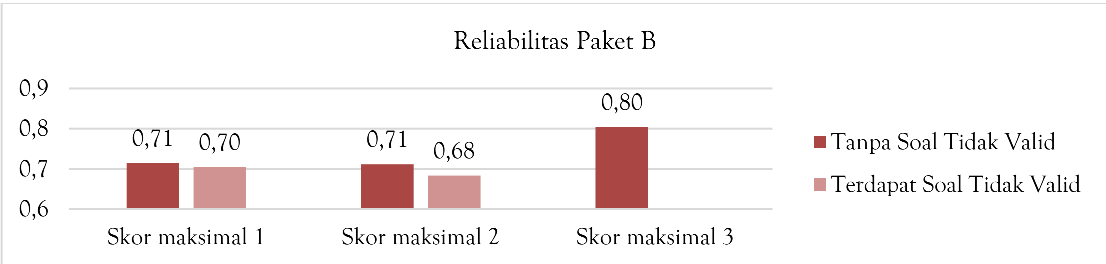


Figure 6. Package B Reliability Bar Chart

The results of the reliability analysis indicate that the reading literacy and numeracy assessment instruments in both packages have high reliability coefficients, thus categorizing them as reliable. This finding indicates that the instruments can provide consistent measurement results when assessing student abilities. This consistency indicates that variations in scores are influenced more by differences in student abilities than by measurement error, making the instruments suitable for use as assessment tools.

The high reliability coefficients are influenced by the systematic instrument development process, starting with the preparation of the questionnaire, content validation by experts, and empirical validity testing before the reliability analysis. This series of processes resulted in test items that were related to the constructs being measured, namely reading literacy, numeracy literacy, and critical thinking skills in stoichiometry. The congruence between items measuring the same construct contributed to the instrument's high internal consistency. This finding aligns with Fatih et al. (2026), who stated that inter-item homogeneity is one of the main factors determining an instrument's reliability coefficient.

The research results also showed that the reliability coefficient value changed when invalid items were included in the calculation. This condition indicates that the quality of each item affects the overall consistency of the instrument. Reliability is determined not only by the number of items, but also by the extent to which

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

each item consistently measures the same construct. Similarly, Malcolm et al. (2019) explained that reliability reflects the consistency of an instrument in producing relatively similar measurements when used on groups with similar characteristics under comparable conditions. The higher the instrument's reliability, the greater the level of confidence in the measurement results and the resulting interpretation of student abilities.

In addition to inter-item homogeneity, instrument reliability is also influenced by several other factors. Defrijon et al. (2025) and Muharromah et al. (2025) explained that an adequate number of items can increase reliability because it can more comprehensively represent the competencies being measured. Conversely, instruments with too few items tend to produce less stable measurements. Furthermore, students' conditions during the test, such as physical health, learning motivation, concentration level, and fatigue, can also influence response consistency, thus impacting the instrument's reliability.

In the context of chemistry learning, particularly in stoichiometry, instruments with high reliability are crucial because they provide consistent measurement results for students' reading literacy and numeracy skills. Consistent assessment results enable teachers to obtain more accurate information about students' conceptual mastery, supporting more precise and accountable decisions about remedial activities, enrichment programs, and learning evaluations.

The results of the difficulty level analysis show that there are variations in the level of difficulty in the reading literacy and numeracy instrument questions developed. The distribution of difficulty levels for each question package is presented in Table 7.

Tabel 7. Distribution of Difficulty Level of Question Items in Package A and B

Category	Number of Questions Items in Package A and Percentage (%)	Number of Questions Items in Package B and Percentage (%)
Easy	5 (20%)	10 (40%)
Medium	17 (68%)	13 (52%)
Difficult	3 (12%)	2 (8%)
Total Questions	25	

Based on Table 6, the majority of questions in Package A (68%) and Package B (52%) fall into the moderate category. These results indicate that the developed instrument has a proportional level of difficulty, effectively measuring students' abilities at various levels of material mastery. The predominance of moderate items is likely due to the systematic development of the instrument based on the AKM framework, indicators of reading literacy, numeracy literacy, critical thinking skills, and learning outcomes in stoichiometry. Therefore, the complexity of each item has been adjusted to the characteristics and abilities of high school students, resulting in a relatively balanced distribution of difficulty levels.

These findings have important implications for chemistry learning, particularly in stoichiometry, which requires the ability to understand information in text form while applying quantitative concepts to solve problems. A predominance of moderate items allows the instrument to measure students' reading literacy and numeracy skills more optimally than if too easy or too difficult questions dominated. Furthermore, a balanced distribution of difficulty levels also helps teachers obtain more accurate information regarding students' conceptual mastery, which can then serve as a basis for designing follow-up learning. According to Evangelista et al. (2022), this condition is likely related to the characteristics of stoichiometry material, which requires not only an understanding of chemical concepts but also the ability to perform mathematical calculations and connect various concepts. Previous research has shown that students still experience difficulties in several stoichiometry subtopics, such as balancing reactions, calculating the mass of substances, and determining moles. Furthermore, most students also experience difficulty performing mathematical operations and require more time to understand stoichiometry problems. This condition indicates that the difficulty level of some questions is influenced not only by the question structure but also by the complexity of the material being measured.

The results of this study align with Arikunto's (2016) opinion, which states that one characteristic of a good instrument is having a balanced distribution of difficulty levels. Similarly, Magdalena et al. (2021) explained that items in the moderate category should be retained in the question bank because they provide more optimal information about student abilities. Therefore, most of the items in the developed instrument have the potential to be reused as assessment instruments for relevant material.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

However, some items still fall into the easy and difficult categories. The presence of both categories is still necessary to measure students' abilities at various levels. However, their proportions must be maintained so that they do not dominate the entire instrument. Yadnyawati (2019) stated that one possible difficulty level composition is a 3:4:3 ratio, namely 30% easy questions, 40% medium questions, and 30% difficult questions. Meanwhile, Fatimah and Alfath (2019), explained that there is no standard for the composition of difficulty levels because it can be adjusted according to the purpose of the test. Therefore, some items that remain in the easy or difficult categories need to be reviewed, particularly regarding wording, question context, and distractor effectiveness, so that the distribution of instrument difficulty levels can be more proportional.

Then, the results of the discriminant power analysis indicate the ability of each question item to differentiate between students with high and low ability. The distribution of discriminating power categories across both question sets is presented in Table 8.

Tabel 8. Distribution of the proportion of the discriminating power of Package A and B test items

Category	Number of Questions Items in Package A and Percentage (%)	Number of Questions Items in Package B and Percentage (%)
Very Good	13 (52%)	11 (44%)
Good	7 (28%)	13 (52%)
Fair	3 (12%)	1 (4%)
Poor	2 (8%)	2 (8%)
Total Questions	25	

Analysis of Table 7 shows that the majority of test items in both packages are classified as good or excellent. In Package A, the very good category accounted for 52% of items, whereas in Package B, the good category comprised 52%. These findings suggest that most test items effectively differentiate students according to their ability levels. This outcome is attributed to the development of each test item using a grid structure aligned with the AKM framework, reading literacy indicators, numeracy literacy, critical thinking skills, and stoichiometry learning outcomes. The systematic structure of the items produced questions with varying complexity, enabling clear differentiation between students who have mastered the concepts and those who continue to experience difficulties.

These results demonstrate that the developed instrument fulfills a key criterion of effective assessment tools: the ability to identify differences in students' conceptual mastery. Discriminatory ability is essential in instrument evaluation, as it reflects how well each test item can group students according to their understanding and mastery of concepts (Rachmawati and Pradana, 2025). In chemistry education, particularly in stoichiometry, this capability is significant because it allows teachers to assess not only the accuracy of students' responses but also their integration of reading and numeracy skills when solving stoichiometry problems.

The average discrimination index of 0.62 in Package A and 0.58 in Package B is considered good. These results indicate that the instrument generally has adequate discriminatory ability and is therefore suitable for measuring student abilities. Good discrimination ability indicates that most test items can consistently differentiate between students with high and low levels of conceptual mastery (Ati et al., 2026). This finding aligns with Pradita et al. (2023), who stated that test items with a low discrimination index, i.e., less than 0.30, are not able to optimally differentiate student abilities. Therefore, although the average discriminatory power of the instrument is in the good category, several items with low discriminatory power still need to be reviewed through improvements to the wording, distractor quality, and indicator suitability to improve the instrument's discriminatory power.

Although the average discriminatory power index is classified as good, some items exhibit low discriminatory power and require further review. Potential causes include ambiguous question wording, ineffective distractors, or inappropriate difficulty levels, which prevent the items from effectively differentiating students by ability. This finding was also reported by Oktaviani et al. (2026), who stated that the quality of discriminatory power is influenced by item construction and the effectiveness of answer alternatives.

Overall, the results of the discriminatory power analysis indicate that the developed reading and numeracy literacy assessment instrument is of good quality, as the majority of items fall into the good and excellent categories. This shows that the instrument is not only able to measure student achievement, but is also effective

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

in differentiating student ability levels so that the assessment results obtained can be a more accurate basis for teachers in designing follow-up learning.

CONCLUSION

The results of the study indicate that the reading and numeracy literacy evaluation instrument for stoichiometry meets the characteristics of a good instrument. Content validation results indicate that most items exhibit a high level of conformity with the measured indicators. Empirical validity testing indicates that most items fall in the moderate-to-high category, and reliability results indicate that the instrument has good internal consistency. In addition, the difficulty level analysis indicates that most of the items are in the moderate category; conversely, the discrimination analysis indicates that most of the items can differentiate students based on their ability levels. Overall, the results indicate that the instrument is suitable for evaluating reading and numeracy skills in stoichiometry. However, some items still require adjustments based on empirical test results to improve the instrument's quality. This study still uses the Classical Test Theory (CTT) approach to evaluate instrument quality. Therefore, further research is recommended to conduct analysis using modern approaches, such as the Rasch Model or Item Response Theory (IRT), so that instrument characteristics can be evaluated more comprehensively.

Bibliography

Adisti Yuliastrin, Deska Indri Yani, & Rian Vebrianto. (2025). Development of Assessment Instruments for Scientific Literacy Skills Tests for VII Grade Junior High School Students on Global Warming Material. *Jurnal Pendidikan Mipa*, *15*(2), 481–490. https://doi.org/10.37630/jpm.v15i2.2576

AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Ambarita, R. S., Wulan, N. S., & Wahyudin, D. (2021). Analisis Kemampuan Membaca Pemahaman pada Siswa Sekolah Dasar. *Edukatif : Jurnal Ilmu Pendidikan*, *3*(5), 2336–2344. https://doi.org/10.31004/edukatif.v3i5.836

Amelia, D., Lu'lua, N., Arafah, N. Q. B., Diaprina, S. R., & Maromy, T. C. (2024). Variabel yang Memengaruhi Kemampuan Literasi Membaca Siswa Indonesia : Analisis Berdasarkan Pendekatan MARS. *Jurnal Pendidikan Dan Kebudayaan*, *9*(2), 205–217. https://doi.org/10.24832/jpnk.v9i2.4966

Amidi. (2024). Literasi Numerasi dalam Pembelajaran Matematika Berbasis Digital. *PRISMA: Prosiding Seminar Nasional Matematika*, *7*, 998–1004.

Arifin, Z. (2012). *Evaluasi Pembelajaran*.

Arikunto, S. (2016). *Dasar-Dasar Evaluasi Pendidikan*. PT Bumi Aksara.

Arini, R., & Rahayu, Y. S. (2023). Profile of Critical Thinking Results Analyzed from Facione Indicators and Gender of Learners. *IJORER : International Journal of Recent Educational Research*, *4*(4), 434–446. https://doi.org/10.46245/ijorer.v4i4.328

Asbupel, F., Retnawati, H., Muhardis, Yovita, S., & Munadi, S. (2021). Analysis of students' difficulties in answering questions on the chemistry national standardized school exam in kupang. *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, *528*. https://doi.org/10.2991/assehr.k.210305.032

Brinus, K. S. W., Makur, A. P., & Nendi, F. (2019). Pengaruh Model Pembelajaran Kontekstual terhadap Pemahaman Konsep Matematika Siswa SMP. *Mosharafa : Jurnal Pendidikan Matematika*, *8*(2). https://doi.org/10.31980/mosharafa.v8i2.558

Budhayanti, C. I. S. (2023). Penggunaan Learning Management System (Lms) Moodle Pada Perkuliahan Pendidikan Matematika Di Program Studi Pendidikan Guru Sekolah Dasar. *Psiko Edukasi*, *21*(1), 50–62. https://doi.org/10.25170/psikoedukasi.v21i1.4439

Cicilia, Y., & Nursalim, N. (2019). Gaya dan Strategi Belajar Bahasa. *Edukatif : Jurnal Ilmu Pendidikan*, *1*(3), 138–149. https://doi.org/10.31004/edukatif.v1i3.30

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Defrijon, Matra, E., Lazuardi, M. R., Kurniadi, Y., Yardi, L., & Julhadi. (2025). RELIABILITAS ALAT UKUR, JENIS-JENIS DAN FAKTOR FAKTOR YANG MEMPENGARUHINYA. *Jurnal Studi Multidisipliner*, *9*(1), 77–86.

Dongoran, F. R., & Hariani, P. P. (2025). Efektivitas Penggunaan Learning Management System (LMS) dalam Pembelajaran Akuntansi : Perspektif Guru dan Siswa. *LIABILITIES: Jurnal Pendidikan Akuntansi*, *8*(1), 63–70. https://doi.org/10.30596/liabilities.v

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). ANALISIS KUALITAS SOAL KEMAMPUAN MEMBEDAKAN RANGKAIAN SERI DAN PARALEL MELALUI TEORI TES KLASIK DAN MODEL RASCH. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19. https://doi.org/10.23887/ijerr.v3i1.24080

Evangelista, E., Ariani, S. R. D., & Hastuti, B. (2022). ANALISIS KESULITAN BELAJAR SISWA KELAS X MIPA DI SMA NEGERI 1 PURWODADI PADA MATERI STOIKIOMETRI DENGAN INSTRUMEN TESLET PADAPEMBELAJARAN JARAK JAUH. *Jurnal Pendidikan Kimia*, *11*(2), 211–220. https://doi.org/10.20961/jpkim.v11i2.63974 211

Fatih, F. Al, Lubis, N. P., Hsb, K. N., Zulpan, & Arianto. (2026). Konsep Homogenitas dan Normalitas dalam Statistik serta Teknik Pengujiannya. *Educational Journal*, *1*(3), 805–817. https://doi.org/10.63822/b98bqe88

Fatimah, L. U., & Alfath, K. (2019a). ANALISIS KESUKARAN SOAL, DAYA PEMBEDA DAN FUNGSI DISTRAKTOR. *Jurnal Komunikasi Dan Pendidikan Islam*, *8*(2), 37–64. https://doi.org/https://doi.org/10.36668/jal.v8i2.115

Fatimah, L. U., & Alfath, K. (2019b). ANALISIS KESUKARAN SOAL, DAYA PEMBEDA DAN FUNGSI DISTRAKTOR. *Al-Manar*, *8*(2), 37–64. https://doi.org/10.36668/jal.v8i2.115

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research In educatIon (8th ed.)* (Vol. 17). Mc Graw HIl.

Geiger, V., & Schmid, M. (2024). A critical turn in numeracy education and practice. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1363566

Gliem, J. A., & Gliem, R. R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *Studies in Inorganic Chemistry*, *14*(C), 349–372. https://doi.org/10.1016/B978-0-444-88933-1.50023-4

Handayani, E. S., Fernando, F., Gaspersz, S., Ridwan, Ahmadin, & Kusumarini, E. (2025). Jurnal Edu Research Indonesian Institute For Corporate Learning And Studies (IICLS) Page 1522. *EDU RESEARCH: Jurnal Penelitian Pendidikan*, *6*(2), 1522–1535. https://doi.org/10.47827/jer.v6i2.975

Hernani, Nurhadi, A. R., Mudzakir, A., Supriatna, A., Wahyuni, S., Depi, S. S., Hidayat, L. N., Irmawati, Fazarwati, R., Aulia, A., Muyassaroh, A., & Fadhilah, A. (2024). Design and Implementation of Contextual Chemistry Learning Oriented towards Science Literacy to Support the Implementation of the Merdeka Curriculum [Perancangan dan Implementasi Pembelajaran Kimia Kontekstual Berorientasi Literasi Sains untuk Mendukung. *Jurnal Pengabdian Isola*, *3*(1), 179–185. https://doi.org/10.17509/jpi.v3i1.70353

Hidayat, A., Rasmawan, R., Oktavianty, E., & Tanjungpura, U. (2024). PENINGKATAN AKTIVITAS DAN HASIL BELAJAR MELALUI PENDEKATAN KIMIA KIMIA KONTEKSTUAL PADA MATERI KONSEP MOL. *JURNAL KAJIAN PEMBELAJARAN DAN KEILMUAN*, *8*(2019), 240–247. https://doi.org/10.26418/jurnalkpk.v8i2.77792

Janah, L., & Kurniawan, A. (2025). Development and Validation of a STEM-Based Minimum Competency Assessment Instrument for Reading Literacy and Numeracy in Stoichiometry. *Panthera : Jurnal Ilmiah Pendidikan Sains Dan Terapan*, *5*(4), 953–966. https://doi.org/10.36312/panthera.v5i4.638

Jayanti, W. P., Nahadi, & Siswaningsih, W. (2026). Development and Validation of a STEM-Based Minimum Competency Assessment Instrument for Reading Literacy and Numeracy in Stoichiometry. *Jurnal*

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Penelitian Pendidikan IPA, *12*(5), 296–305. https://doi.org/10.29303/jppipa.v12i5.14571

Kamaludin. (2024). MENDEKONSTRUKSI KEMEROSOTAN PERILAKU PESERTA DIDIK INDONESIA : STUDI KRISIS LITERASI , NUMERASI , DAN DISIPLIN DALAM BAYANG-BAYANG. *Transformasi KEMENAG*, *7*(1), 132–144.

Karira, N. F., Sunarti, T., J, M. N. R., & Setyasih, W. (2023). Validitas Instrumen Tes Berbasis Literasi Sains untuk Mengukur Keterampilan Berpikir Kritis Siswa SMA pada Materi Energi Terbarukan. *Inovasi Pendidikan Fisika*, *12*(2), 118–125. https://doi.org/10.26740/ipf.v12n2.p118-125

Khan, H. F., Qayyum, S., Beenish, H., Khan, R. A., & Iltaf, S. (2025). Determining the alignment of assessment items with curriculum goals through document analysis by addressing identified item flaws. *BMC Medical Education*. https://doi.org/10.1186/s12909-025-06736-4

Lawshe, C. H. (1975). A QUANTITATIVE APPROACH TO CONTENT VALIDITY. *Pers. Psychol.*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). KESULITAN DAN DAYA BEDA BUTIR SOAL UJIAN AKHIR SEMESTER TEMA 7 KELAS III SDN KARET 1 SEPATAN. *BINTANG : Jurnal Pendidikan Dan Sains*, *3*(2), 198–214. https://doi.org/10.36088/bintang.v3i2.1291

Malcolm, S. A., Mavhunga, E., & Rollnick, M. (2019). The Validity and Reliability of an Instrument to Measure Physical Science Teachers' Topic Specific Pedagogical Content Knowledge in Stoichiometry. *African Journal of Research in Mathematics, Science and Technology Education*, *23*(2), 181–194. https://doi.org/10.1080/18117295.2019.1633080

Metafisika, K. (2018). Analisis Ketepatan Deskripsi Konsep pada Buku Teks Kimia SMA pada materi. *Jurnal Konfigurasi*, *1*(1), 11–22. https://doi.org/10.24014/konfigurasi.v2i1.4512

Muharromah, F., Fani, F. Al, & Ridho, A. R. (2025). RELIABILITAS ALAT UKUR: JENIS-JENIS, CARA PENGUKURAN, DAN FAKTOR FAKTOR YANG MEMPENGARUHINYA DALAM PEMBELAJARAN. *JIPDAS : Jurnal Ilmiah Pendidikan Dasar*, *4*(1), 63–73. https://doi.org/10.65946/8qbjfd05

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. https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i3.33067

Novita, N., Mellyzar, & Herizal. (2021). Asesmen Nasional (AN): Pengetahuan dan Persepsi Calon Guru. *JISIP: Jurnal Ilmu Sosial Dan Pendidikan*, *5*(1). https://doi.org/10.58258/jisip.v5i1.1568

OECD. (2022). *PISA 2022 Results (Volume 1): The State of Learning and Equity in Education*.

Oktaviani, C., Mauliza, M., & Nurhafidhah, N. (2026). Development of E-Assessment to Measure Critical Thinking in Chemistry Learning. *EduChemia: Jurnal Kimia Dan Pendidikan*, *11*(1). https://doi.org/10.30870/educhemia.v11i1.36935

Parhan, M., & Sutedja, B. (2019). PENERAPAN PENDEKATAN PEMBELAJARAN KONTEKSTUAL DALAM PENDIDIKAN AGAMA ISLAM DI UNIVERSITAS PENDIDIKAN INDONESIA. *TARBAWY: Indonesian Journal of Islamic Education*, *6*(2), 114–126. https://doi.org/10.17509/t.v6vi2.20165

Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran , Daya Pembeda , dan Fungsi Distraktor. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, *3*(80), 109–118.

Pratama, C., Dimas Julian Pratama, Rakha Agusti Amanullah, Sani Safitri, & Rani Oktapiani. (2025). Relevansi Macam-Macam Pendekatan Pembelajaran Abad 21 terhadap Kebutuhan Siswa di Era Revolusi Digital. *SOSIAL: Jurnal Ilmiah Pendidikan IPS*, *3*(4), 198–209. https://doi.org/10.62383/sosial.v3i4.1362

Puspitasari, E., Abdullah, A., & Wati, M. (2025). Validasi: LKPD Terintegrasi Model Problem Based Learning untuk Melatih Keterampilan Berpikir Kritis Materi Asam Basa. *Jurnal Pendidikan Mipa*, *15*(1), 129–137.

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

https://doi.org/10.37630/jpm.v15i1.2405

Rachmawati, D., & Pradana, A. B. (2025). Analisis butir soal mata pelajaran ekonomi: validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. *Jurnal Pendidikan Ekonomi (JUPE)*, *13*(3), 273–284. https://doi.org/10.26740/jupe.v13n3.p273-284

Rahmawan, S., & Ashfarini, D. (2024). Identification of Misconceptions and Causes of Misconceptions in Stoichiometry Material Using Four Tier Multiple Choice (4TMC). *Orbital: Jurnal Pendidikan Kimia*, *8*(1), 39–48. https://doi.org/10.19109/ojpk.v8i1.21649

Risti, S. (2025). Sistem E-Learning Berbasis Moodle. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, *4*(1), 79–85. https://doi.org/10.62712/juktisi.v4i1.358

Rohim, D. C., Rahmawati, S., & Ganestri, I. D. (2021). Konsep Asesmen Kompetensi Minimum Meningkatkan Kemampuan Literasi Numerasi Sekolah Dasar untuk Siswa. *Jurnal Varidika*, *33*(1), 54–62. https://doi.org/10.23917/varidika.v33i1.14993

Rohman, A., Islam, U., & Sunan, N. (2022). Literasi dalam Meningkatkan Kemampuan Berpikir Kritis di Era Disrupsi. *EUNOIA: Jurnal Pendidikan Bahasa Indonesia*, *2*(1), 40–47. https://doi.org/10.30821/eunoia.v2i1.1318

Salim, A., & Juita, N. (2025). Analisis Pentingnya Pemanfaatan dari Instrumen Penelitian. *AL-AFKAR: Journal for Islamic Studies*, *8*(4), 880–888. https://doi.org/10.31943/afkarjournal.v8i4.1700.Analysis

Saputri, H. A., Zulhijrah, Larasati, N. J., & Shaleh. (2023). ANALISIS INSTRUMEN ASSESMEN : VALIDITAS, RELIABILITAS, TINGKAT KESUKARAN DAN DAYA BEDA BUTIR SOAL. *Didaktik : Jurnal Ilmiah PGSD FKIP Universitas Mandiri*, 09(05), 2986–2995. https://doi.org/10.36989/didaktik.v9i5.2268

Sari, F. M., Hadiati, R. N., & Sihotang, W. P. (2023). Analisis Korelasi Pearson Jumlah Penduduk dengan Jumlah Kendaraan Bermotor di Provinsi Jambi. *Multi Proximity: Jurnal Statistika Universitas Jambi*, *2*(1), 39–44.

Selaras, G. H., Ahda, Y., & Alberida, H. (2019). The Validity and Reliability of The Instrument Assessment of Higher Order Thinking Skill on The Biological Scope Materials. *Bioeducation Journal*, *3*(2), 124–125. https://doi.org/10.24036/bioedu.v3i2.242

Setiyawan, R. A., & Wijayanti, P. S. (2020). Analisis kualitas instrumen untuk mengukur kemampuan pemecahan masalah siswa selama pembelajaran daring di masa pandemi. *Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *1*(2), 130–139. https://doi.org/10.46306/lb.v1i2 ANALISIS

Skills, P. for 21st C. (2019). *PARTNERSHIP FOR 21ST CENTURY SKILLS- Framework for 21st Century Learning*.

Sudiana, R. (2016). EFEKTIFITAS PENGGUNAAN LEARNING MANAGEMENT SYSTEM BERBASIS ONLINE. *JPPM: Jurnal Penelitian Dan Pembelajaran Matematika*, *9*(2), 201–209. https://doi.org/10.30870/jppm.v9i2.998

Sugiono, Noerdjanah, & Wahyu, A. (2020). Uji Validitas dan Reliabilitas Alat Ukur SG Posture Evaluation. *Poltekkes Surakarta: Jurnal Keterapian Fisik*, *5*(1), 55–61. https://doi.org/10.37341/jkf.v5i1.167

Sumarno. (2020). ANALISIS ISI DALAM PENELITIAN PEMBELAJARAN BAHASA DAN SASTRA. *Edukasi: Lingua SantrA*, *18*(2). https://doi.org/10.47637/elsa.v18i2.299

Utami, N. D. (2024). Penerapan Pendekatan Kontekstual untuk Meningkatkan Kemampuan Pemecahan Masalah Matematis Siswa Sekolah Dasar. *PGSD: Jurnal Pendidikan Guru Sekolah Dasar*, *1*(2), 1–9. https://doi.org/10.47134/pgsd.v1i2.134

Utami, S. R., & Kusmayati, N. B. (2024). PENGENALAN LITERASI DAN NUMERASI PADA GURU-GURU BAHASA INDONESIA DAN NON-EKSAKTA SMP. *Jurnal Pengabdian Kepada Masyarakat*, *13*(1), 44–51. https://doi.org/https://doi.org/10.23887/jwl.v13i1.69055

Wirati, N. N. (2023). Penerapan Model Pembelajaran Contextual Teaching and Learning dalam Upaya

ISSN: 2088-0294 | e-ISSN: 2621-9166
https://doi.org/10.37630/jpm.v16i4.4950

Meningkatkan Aktivitas dan Prestasi Belajar Siswa. *Journal of Education Action Research*, *7*(4), 508–517. https://doi.org/10.23887/jear.v7i4.66558 Penerapan

Wulan. (2023). ANALISIS CAPAIAN PEMBELAJARAN KURIKULUM MERDEKA PADA MATERI ASAM BASA. *TSAQOFAH: Jurnal Penelitian Guru Indonesia*, *4*, 58–65. https://doi.org/https://doi.org/10.58578/tsaqofah.v4i1.2129

Yadnyawati, I. A. G. (2019). *Evaluasi Pembelajaran* (I. K. Suda (ed.)). UNHI PRESS 2019.

Yustiqvar, M., Hadisaputra, S., & Gunawan, G. (2019). ANALYSIS OF STUDENT CONCEPTS MASTERY FOR THE INTERACTIVE MULTIMEDIA BASED GREEN CHEMISTRY APPLICATION. *Jurnal Pijar MIPA*, *14*(3), 135–140. https://doi.org/10.29303/jpm.v14i2.1299