Vol 15 (2), 2026, 202-217 DOI: 10.23960/jppk.v15.i2.15217
Jurnal Pendidikan dan Pembelajaran Kimia
e-ISSN: 2714-9595| p-ISSN 2302-1772
http://jurnal.fkip.unila.ac.id/index.php/JPK/index

Feasibility of Integrated AKM Instrument for Problem Solving on Reaction Rate Factors

Dinda Putri Mugiani, Wiwi Siswaningsih*, Febriati Dian Mubarokah

Chemistry Education, Department of Chemistry Education, Faculty of Mathematics and Natural Sciences Education (FPMIPA), Indonesia University of Education.
Jl. Dr. Setiabudhi No. 229, Sukasari, Isola, Bandung, Bandung City, West Java, Indonesia
.
*Correspondinge-mail: wiwisiswaningsih@upi.edu

Received: July 11th, 2026 Accepted: July 25th, 2026 Online Published: August 28th, 2026

Abstract: Feasibility of Integrated AKM Instrument for Problem Solving on Reaction Rate Factors. This study aims to produce an integrated Minimum Competency Assessment (AKM) instrument for problem-solving skills in the sub-material of factors that influence reaction rate that are worthy of being reviewed in terms of validity and reliability. The method used is development and validation. The participants involved were 4 lecturers and 1 teachers as validators and 35 grade XI students at one of Bandung's public high schools as respondents. The resulting product consists of 25 reading literacy questions and 5 numeracy literacy questions. The product is declared valid based on content validity, while empirically shows 20 valid questions and 10 invalid questions. All questions are declared reliable with details of reliability values for questions with a maximum score of 1 being 0.59, questions with a maximum score of 2 being 0.75, and for questions with a maximum score of 4 being 0.82.

Keywords: Minimum Competency Assessment (AKM), Reading Literacy, Numeracy Literacy, Reaction Rate

Abstrak: Kelayakan Instrumen AKM Terintegrasi Penyelesaian Masalah pada Faktor Laju Reaksi. *Penelitian ini bertujuan untuk menghasilkan instrumen Asesmen Kompetensi Minimum (AKM) terintegrasi kemampuan problem solving pada sub materi faktor-faktor yang memengaruhi laju reaksi yang layak ditinjau dari segi validitas dan reliabilitas. Metode yang digunakan adalah pengembangan dan validasi. Partisipan yang terlibat yaitu 4 dosen dan 1 guru sebagai validator serta 35 murid kelas XI di salah satu SMA Negeri Bandung sebagai responden. Produk yang dihasilkan terdiri dari 25 soal literasi membaca dan 5 soal literasi numerasi. Produk dinyatakan valid berdasarkan validitas isi, sedangkan empiris menunjukkan 20 soal valid dan 10 soal tidak valid. Seluruh soal dinyatakan reliabel dengan rincian nilai reliabilitas untuk soal dengan maksimal skor 1 adalah 0,59, soal dengan skor maksimal 2 adalah 0,75, dan untuk soal dengan skor maksimal 4 adalah 0,82.*

Kata kunci: *Asesmen Kompetensi Minimum (AKM), Literasi Membaca, Literasi Numerasi, Laju Reaksi*

This is an open access article under the CC BY-SA license

▪ INTRODUCTION

In the 21st century, technological developments, globalization and the needs of the world of work are occurring rapidly (Rismana & Hernawati, 2025). This century is often referred to as the century of knowledge, industrial revolution 4.0, knowledge-based economy, and so on (Rosnaeni, 2021). The developments that occur require adaptation to

an education system that not only has mastery of intellectual knowledge, but includes 21st century educational competencies which include critical thinking, problem solving abilities, the ability to innovate and be creative, communicate, collaborate, and adapt to technological advances(Halim, 2022). Basic competencies that are of concern to realize 21st century skills such as students' ability to understand information, interpret data, and make decisions rationally. This refers to reading literacy and numeracy skill (Fahmy et al., 2025). Based on data from the Ministry of Education and Culture's Education Report Card, students who achieved the minimum competency in reading literacy nationally experienced an increase from 59.49% in 2022, to 68.05% in 2023, and increased again to 70.03% in 2024. Meanwhile, students who achieved the minimum competency in numeracy also experienced an increase from 45.24% in 2022, to 62.45% in 2023, and to 67.94% in 2024. Although the results show an increase, the data confirms that the increase or achievement of this minimum competency is not evenly distributed across all districts/cities in Indonesia (Kementerian Pendidikan Dasar dan Menengah, 2025). To overcome this, the government has begun to introduce the Minimum Competency Assessment (AKM) as part of the National Assessment (AN). Previously, students took the National Examination (UN), which was then officially abolished in 2021 and replaced with the National Assessment. The National Assessment includes the Minimum Competency Assessment (AKM) which aims to measure basic student competencies, namely reading literacy and numeracy skills as part of educational evaluation (Inayah et al., 2022).

One of the subjects that has demands on aspects of reading literacy and numeracy is chemistry. Each chemical material has its own characteristics so that it can improve students' reading literacy and numeracy(Rosmiati, 2022). This is because chemistry content and sub-material require students to think critically in solving a problem, chemistry content can be in the form of reading or calculations (Rokhim et al., 2022). Based on research conducted by (Daeli et al., 2025), students tend to have difficulty understanding reading and interpreting data in the form of tables or graphs. This condition shows that learning and assessment that focuses on memorization has not been fully capable of obtaining the expected minimum competency results. Many students have difficulty understanding chemistry concepts because of their abstract nature, which has a significant impact on students' low academic achievement. One material that often triggers misconceptions (inappropriate understanding) is the reaction rate, especially in the sub-material factors that influence the reaction rate. Research conducted by Lestari (2021) shows clear evidence that there are still high misconceptions about this material with data that 22.58% of students experienced misconceptions about the concentration factor, 16.13% about the surface area factor, 38.71% about the temperature factor, and 12.90% about the catalyst factor. This ongoing challenge is reflected in the continued research attention paid to this topic, with recent studies such as that of (Lubay et al., 2024) specifically examining students' analytical skills on the topic of reaction rates, indicating that this sub-topic remains an active area of instructional research. In practice, students often just memorize without understanding its application in everyday life. This indicates that learning and assessment that still focuses on memorization has not been able to build a deep understanding of concepts in students.

Therefore, it is necessary to have an assessment approach that does not only measure memorization, but can measure students' reading literacy and numeracy skills in more depth. In line with AKM, the Ministry of Education and Culture introduced a deep learning approach or Deep Learning as an innovation in the 21st century learning

paradigm. Deep Learning is a learning approach that emphasizes a conscious, meaningful and joyful learning process. One of the main competencies in Deep Learning is critical thinking and problem solving skills which can encourage students to look for innovative solutions(Handayani, 2025). Research conducted by Hidayatulloh & Utiya Azizah (2020) shows that problem solving can be integrated into learning and is in line with the deep learning approach which emphasizes deep understanding and high-level reasoning. This approach is relevant to the demands of AKM which measures literacy and numeracy skills in the context of problem solving. In the AKM context, problem solving does not stand as a separate ability, but is integrated in students' cognitive processes when solving reading literacy and numeracy questions (Handayani, 2025).

However, AKM instruments integrated with problem-solving skills in chemistry, particularly the sub-topic of factors influencing reaction rates, are still limited. Several previous studies have developed assessment instruments for reaction rates, including the study by Zahroh (2019), which developed a problem-solving ability-based instrument for reaction rate. However, this study did not yet utilize the AKM framework. Another study conducted by Rosyada et al. in 2023 developed HOTS-integrated AKM questions for general reaction rate material. This study also found that teachers still rarely implement AKM questions. However, this study did not specifically explore the factors influencing reaction rates that integrate problem-solving skills.

Although the two studies have contributed to the question instrument, no specific study has been found that integrates the AKM framework with problem-solving skills in the sub-topic of factors influencing reaction rates, a sub-topic that, according to Lestari's findings at 2021, still has high levels of misconceptions. Zahroh's study used problem-solving skills but did not use the AKM framework, while Rosyada's study did not target specific sub-topics and did not explicitly integrate problem-solving. Both studies also failed to present concrete contexts such as sustainability issues (SDGs) or the use of digital platforms. This gap in the integration of the AKM framework, problem-solving, specific sub-topics, the SDG context, and the use of digital platforms is the basis for the urgency and novelty of this study.

In addition to integrating problem-solving skills, to make the assessment instruments more meaningful, contexts relevant to everyday life are also needed. One example is the use of SDG Goal 12, which is closely related to the concept of reaction rate, namely "ensuring sustainable consumption and production patterns," as the context for the stimulus questions (Capah et al., 2023). The relationship between the concept of reaction rate and SDGs goal 12, such as catalysts used in industrial processes that allow reactions to occur faster even at low temperatures, thereby reducing energy consumption and emissions in the production process. Furthermore, understanding the factors of concentration and surface area also plays a crucial role in optimizing the use of raw materials for greater efficiency, thereby minimizing production yields. Thus, the sub-topic on factors affecting reaction rate is conceptually relevant to SDG 12 and provides a meaningful context for students in understanding the relationship between chemistry and sustainability issues in everyday life. Furthermore, technological developments have also driven significant changes in the implementation of educational assessments, presenting them in more interactive and contextual formats through digital platforms (Bankole & Ayanwale, 2025).

Based on this description, the development of contextualized AKM instruments that integrate 21st-century skills in line with current educational demands, one of which is problem-solving skills, is still needed. Therefore, this study was conducted to develop

an AKM instrument to measure reading literacy and numeracy skills by integrating problem-solving skills, using the context of SDG Goal 12, and conducting it through a digital platform, specifically for the sub-topic on factors influencing reaction rate.

▪ METHOD

The research design used in this study is the "Development and Validation" method, based on the method developed by Adams and Wieman, which aims to produce assessment instruments that are content-valid and empirically reliable (Adams & Wieman, 2010). This research method is implemented through the following stages:

1. Planning Stage

This stage involved an analysis of learning outcomes in phase F in the merdeka curriculum on reaction rates, focusing on factors influencing reaction rates. It was supported by a review of high school and general chemistry textbooks, as well as other relevant sources related to the topic. Reviewing the AKM framework for reading literacy and numeracy was reviewed as the basic for developing items aligned with established standards with the integration of problem-solving skills and contextualized the scope of SDG goal 12. (Amali et al., 2024)

2. Development Stage

The development of the AKM instrument began by mapping the Phase F learning outcomes on reaction rates into specific indicators, followed by determining the reading text title, subtopic, reading literacy and numeracy aspects according to the AKM framework, item indicators, question format, and question number, consistent with the development procedures applied in previous chemistry instrument development research(Suwahono et al., 2025). Based on this framework, 40 items were developed in various formats, including multiple choice, complex multiple choice, essay, short answer, and matching, accompanied by reading texts designed to reflect problem-solving skills. Content validation was conducted using the AKM instrument validation sheet; items that did not meet the validators' criteria were revised according to their suggestions, particularly regarding clarity of language and alignment between the reading stimulus and the questions, and then could be used in the pilot test. Of the 40 validated items, 30 were selected for the pilot test phase, taking into account the time required to complete the reading literacy and numeracy items, as the AKM and problem-solving aspects measured by the other 10 items were adequately represented by the selected items.

3. Trial and Data Analysis Stage

Questions declared valid by the validator were piloted on a limited basis with 11th-grade students to collect data. The data obtained from the test were then analyzed and calculations were performed to test empirical validity and reliability. All tests were conducted to identify items that met the requirements for a good AKM instrument. (Amali et al., 2024)

4. Evaluation Stage

The evaluation stage was conducted to ensure the achievement of the research objectives and to confirm the quality of the developed instrument.

Participants and Research Site

The participants in this study were five validators participated: 4 chemistry lecturers from the Faculty of Mathematics and Natural Sciences, Universitas Pendidikan Indonesia (UPI) and 1 chemistry teacher from the high school where the study was conducted, and 35 students in grades XI-4 as a respondent. The validity of the research instrument was examined through expert assessment of the content and empirical results of student answers. Initially, a 40-item instrument was developed, of which 30 items were selected for the pilot phase, considering the efficiency of the instrument's time while maintaining the representation of each indicator. The instrument consists of various question formats including multiple choice, complex multiple choice, essay, fill-in-the-blank, and matching items, designed to measure students' reading literacy and numeracy skills integrated with problem-solving abilities.

Data Analysis techniques

Data analysis techniques included validity and reliability tests to assess the feasibility of the AKM question instruments developed using Microsoft Excel software.

Validity test

According toLawshe (1975), there are stages carried out to obtain content validation data is calculated using the CVR (Content validity ratio) with the equation:

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

Description:
n_e= number of validators who stated essential
N = total number of validator
With 5 validators, the question is declared valid in terms of content when the CVR value is >0.99.

An empirical validity test formula used is:

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

Description:
rxy = desired correlation coefficient
N = number of test participants
X = value of variable X (item score)
Y = value of variable Y (item score)
If r count > r table, the item is declared valid. (Riyani & Maizora, 2017)

There are several criteria for the validity of the test items obtained. The following table presents the categories for the product-moment correlation coefficient.

Table 1. Validity Level Categories for Item Test

Correlation Coefficient	Category
0,80< r_{xy} ≤1,00	Very high
0,60< r_{xy} ≤0,80	High
0,40< r_{xy} ≤0,60	Fair
0,20< r_{xy} ≤0,40	Low
0,00< r_{xy} ≤0,20	Very low

(Son, 2019)

The categories proposed by Son reflect the strength of the relationship between item scores and the total test score. In general, items with very high and high correlation coefficients indicate that the item is highly representative of the construct being measured. Items with moderate correlation coefficients indicate a moderate relationship and provide adequate evidence of construct representation. Meanwhile, items with low and very low correlation coefficients indicate that the item contributes less to measuring the intended construct.

Reliability test

To measure the reliability of multiple-choice and short answer questions with a maximum score of one (dichotomous), the Kuder-Richardson 20 (KR-20) technique can be determined using the formula:

$$\rho KR20 = \frac{k}{k-1}\left(1 - \frac{\sum pq}{\sigma 2}\right)$$

Description:
ρKR20 = Reliability coefficient
k = Number of items
p = Proportion of correct answers
q = Proportion of incorrect answers
σ2 = Test score variance

The reliability values obtained will then be categorized according to the following criteria by Islamiyah & Yuliani (2024) :

Table 2. KR-20 Reliability Criteria KR-20

Value Range	Category
0,81 - 1,00	Very reliable
0,61 - 0,80	Reliable
0,41 – 0,60	Quite reliable
0,21 – 0,40	Less reliable
<0,20	Not reliable

Adapted on the reliability categories proposed by (Farida & Musyarofah, 2021), the very reliable category reflects a very high level of consistency in measurement results. Reliable indicates good consistency. Moderately reliable indicates moderate consistency. Less reliable reflects relatively low consistency, while unreliable indicates inconsistent measurement results.

To measure the reliability of multiple-answer, matching, and essay questions with a maximum score of more than one (polyatomic), Cronbach's Alpha technique can be used with the formula:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum S_i^2}{S_t^2}\right)$$

Where :
α = Item reliability
k = Number of items
S_i^2 = Variance per item
S_t^2 = Total score variance

The level of item reliability depends on the correlation coefficient obtained. The reliability categories by Gliem & Gliem are presented in the following table:

Table 3. Cronbach's Alpha reliability criteria

Value Range	Category
$\alpha \geq 90$	Very high
$0{,}80 \leq \alpha < 0{,}90$	High
$0{,}70 \leq \alpha < 0{,}80$	Acceptable
$0{,}60 \leq \alpha < 0{,}70$	Doubtful
$0{,}50 \leq \alpha < 0{,}60$	Poor
$\alpha < 0{,}50$	Unacceptable

Each category reflects the level of internal consistency among items, ranging from very high ($\alpha < 0.90$), indicating the items consistently measure the construct, to unacceptable ($\alpha < 0.50$), indicating the items do not consistently measure the same construct. A higher Cronbach's alpha coefficient indicates greater internal consistency, meaning the instrument measures the intended construct more consistently (Gliem & Gliem2003). The overalll instrument was considered reliable and suitable for use if the reliability coefficienet for each item reached at least the quite reliable (KR-20) or acceptable (Cronbach's Alpha).

▪ RESULT AND DISCUSSION

The development of an integrated minimum competency assessment instrument for problem-solving skills in the context of SDGs goal 12 goes through 4 stages, namely: planning, development, testing and data analysis, and evaluation.

Planning stage

The planning phase generated a basic framework for instrument development, derived from an analysis of learning outcomes in Phase F of the Independent Curriculum on reaction rates. This was supported by a review of secondary and general chemistry textbooks and other relevant sources. Furthermore, a review of the 2023 AKM framework from the Pusat Asesmen Pendidikan (Pusmendik, 2023) served as a reference for developing test items to ensure alignment with applicable standards, which were then integrated with problem-solving skills and used in the context of SDG 12.

Development stage

Based on the results of the planning stage, a grid of AKM instruments was prepared which included the title of the reading text, learning outcomes, reaction rate subtopics, aspects of reading literacy and numeracy according to the AKM framework, item indicators, question formats, and question numbers. From this grid, 40 questions were developed in various formats, namely multiple choice, complex multiple choice, essay, fill-in-the-blank, and matching, with stimuli integrating the context of SDGs goal 12, as presented in Table 4.

Tabel 4. Example of an AKM instrument grid

Learning Outcomes	**Sub-topic**	**Question Item Indicators**	**Aspect of reading literacy and numeracy literacy**	**Problem Solving Indicator**	**Question Format**	**Question Number**
Students can analyze factors that influence reaction rate.	The effect of temperature on reaction rate	Analyze factors that influence reaction rate based on the text about the phenomenon of apple browning presented.	**Reading Literacy** •Content: Informational Text •Context: Scientific •Competency: Finding Information •Sub-competency: Searching for and selecting relevant information • Sub-competency Details: Identifying effective keywords to find relevant information sources in informational texts that increase in level	-Aspect: Identifying problems -Indicator Details: Determining concepts or categories	Multiple choice	2
	The effect of concentration on reaction rates.	Analyze the effect of different concentrations of hypochlorite in bleach on the time required to remove color from fabric by	**Numeracy Literacy** • Content: Informational text • Context: Scientific • Cognitive level: Reasoning	Aspect: Examining the problem Indicator details: Examining the interrelationships of the given problem.	Multiple choice	37

Learning Outcomes	Sub-topic	Question Item Indicators	Aspect of reading literacy and numeracy literacy	Problem Solving Indicator	Question Format	Question Number
		interpreting data in the table.	• Domain: Data and Uncertainty •Subdomain: Data and its representation			

Based on the prepared grid, instrument questions were then created. Each question item was structured to measure the indicators set out in the grid. Forty questions were developed, including multiple choice, complex multiple choice, fill-in-the-blank, descriptive, and matching. An assessment rubric was also created to facilitate the data processing process of the instrument trial results and ensure consistency in the assessment process. For multiple choice, multiple answer with two correct answers, and short answer questions, the assessment rubric contained scoring criteria based on the accuracy of the student's answer. Answers that matched the answer key were given a score of 1, while answers that did not match were given a score of 0. For matching questions, the assessment rubric was designed with a scoring range of 0 to 4. For descriptive questions, the maximum score given if the student's answer matched all the predetermined keywords was 2, if only partially met the keywords then the score obtained was 1, and if none of the keywords were met then the score obtained was 0.

After the initial draft of the instrument was completed, content validity was carried out by 4 expert lecturers and 1 chemistry teacher as validators. The assessment was then calculated using the Content Validity Ratio (CVR) method according to Lawshe 1975 to determine whether a question item was valid or not according to experts with 3 assessment aspects, namely the suitability of the text to the question item, the suitability of the question item indicator to the question item, and the suitability of the problem solving indicator to the question item, the following results were obtained:

Table 5. Overall CVR Value

Content Validity Aspect	Number of Valid Question Items Based on CVR	Description
Correspondence of Text to Question Items	39	Question Item 39 is Invalid
Correspondence of Question Item Indicators to Question Item	39	Question Item 39 is Invalid
Correspondence of Problem Solving Indicators to Question Item	39	Question Item 39 is Invalid

Based on the CVR value calculation data, 39 valid questions were obtained based on the content because they had a CVR value >0.99, so these questions can be used for the trial stage for students. However, one question was declared invalid because the

stimulus used in the question is still general and not contextual, so it needs to be revised first. This result means that the developed test instrument is in line with the aspects of reading literacy and numeracy as well as the objectives that have been previously formulated.

Testing and Data Analysis stage

After the instrument was declared suitable for use, the third stage was carried out, namely the trial and data analysis stage. The trial stage was carried out on 35 11th grade students at a public school in Bandung who had studied the material on reaction rates, especially the sub-material on factors that influence reaction rates. Data obtained from the results of the field trial were then analyzed to assess the quality of the test items including empirical validity and reliability integrated with problem-solving skills. After all stages of instrument development were carried out, it can be concluded that the instrument development process had been carried out systematically in accordance with the development and validation method according to Adam & Wieman (2010).

The results of the trial will yield data in the form of scores, which will then be correlated to determine the empirical validity of each question item. Empirical validity is calculated using the Pearson's Product Moment Correlation Coefficient (r-count). A question is considered valid if the r-count is greater than the r-table. The r-table value for a sample of 35 students is 0.334, with a significance level of 0.05. The results of the empirical validity analysis of the instruments that have been tested on students are presented in **Figure 1.**

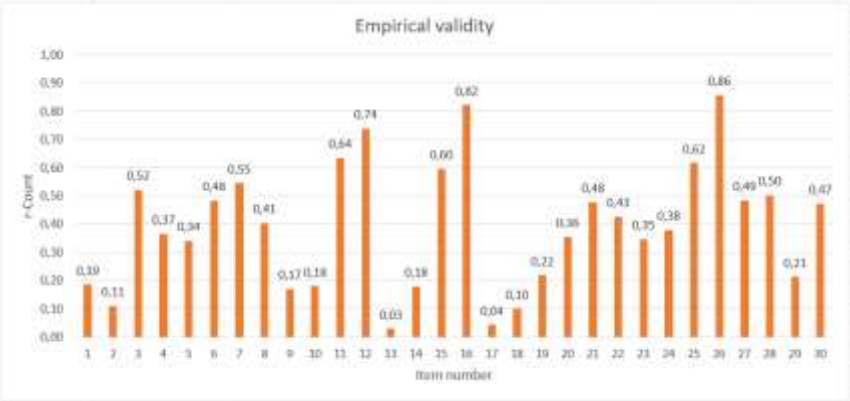


Figure 1. Bar chart of empirical validity of each question item

Based on Figure 1, of the 30 questions tested on students, 10 items were declared invalid (r-count < 0.334), of which 8 items fell into the very low category and 2 items into the low category. It should be noted that the invalid items were predominantly multiple-choice items with a dichotomous scoring scheme, indicating that the item format itself contributed to the invalidity pattern. For example, item number 13, which requires students to identify a problem to be addressed based on a given discourse text, and item number 17, which asks students to judge whether a statement is true or false based on a given discourse text and their reasons, both fell into the very low validity category. Both items have a true/false or single-answer structure that allows a higher probability of correct answers by guessing, thus possibly reducing the correlation between item scores and total scores. In contrast, item number 26, a matching item with a maximum score of four, obtained the highest validity value among all items. This format requires students to correctly pair multiple elements together, making the correct answer much more difficult

to obtain through guesswork and thus resulting in a stronger correlation with the total score. In accordance with Son (2019), there were 8 questions in the very low category, 7 questions in the low category, 10 questions in the fair category, 3 questions in the high category, and 2 questions in the very high category.

The invalidity of several test items in this study is thought to be influenced by several factors. According to Nahadi & Firman (2019),the main factors that can affect validity scores are factors in the text itself, such as incorrect wording and overly complex sentences that confuse students. Furthermore, there are factors influenced by students' responses, such as giving random answers or simply guessing on multiple-choice questions (Supiati & Mujib, 2025). Furthermore, questions that are too easy cause most students to give the same answer, resulting in small score variations. This condition results in a decrease in the correlation between item scores and the total score, so the test items do not meet empirical validity criteria. This pattern is similar to the findings reported by Purniasari et al., (2021) who found that 20% of classically analyzed chemistry test items were declared invalid, largely due to similar item formats and difficulty distributions. Compared with previous research, the proportion of invalid items in this study (33.3%) was relatively higher, possibly due to the more complex cognitive demands. Meanwhile, the matching item, produced the highest validity value. This is in line with the discrimination and reliability patterns reported by Irmayta et al., (2017) in the development of an Arrhenius acid-base knowledge assessment instrument, where items with structured and multi-element response formats consistently achieved high discrimination and reliability categories.

The next test conducted is a reliability test. Reliability testing is conducted to determine how consistently an instrument performs its measurements. Therefore, an instrument is considered good if it is reliable, consistent, and stable, meaning that it will still produce the same data when used repeatedly (Saputri et al., 2023). The higher the reliability score, the more consistent the measurement results, and the more reliable the instrument.

In this study, two techniques were used to determine the reliability of an instrument. This was done based on the type of questions and the total score. The techniques used were Kuder-Richardson 20 (KR-20) and Cronbach's Alpha. The KR-20 technique is used to determine the reliability of items that have a maximum score of 1 (dichotomous) including items number 1, 2, 3, 4, 5, 6, 7, 9, 10, 13, 14, 17, 18, 19, 22, 23, 24, 27, and 29. Then the Cronbach's Alpha technique is used to determine the reliability of items that have a maximum score of more than one (polyatomic) including items number 8, 11, 12, 15, 16, 20, 21, 25, 26, 28, and 30. The following diagram displays the reliability value for each question item.

The KR-20 technique has a formula that systematically calculates the proportion of correct and incorrect answers, so it tends to be used on questions with a maximum score of 1 and two answer choices: correct and incorrect. Meanwhile, the Cronbach's alpha calculation uses the total variance of the test items, not just focusing on correct and incorrect answers. Therefore, this technique tends to be used on questions with a value range, such as 0-4 or 1-5, and so on (Subhaktiyasa, 2024).

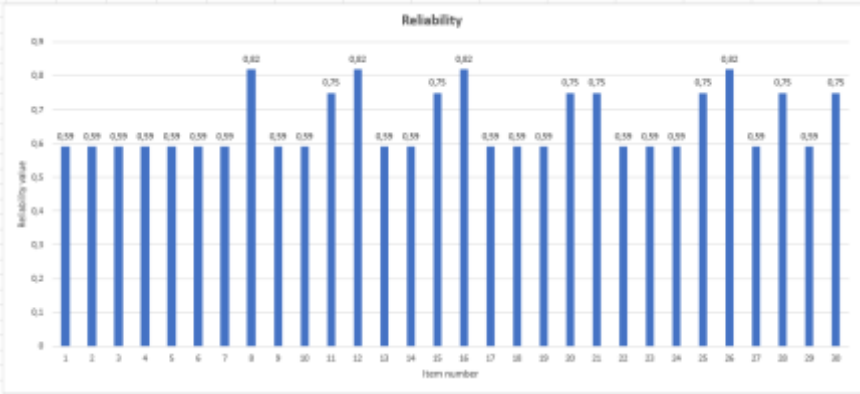


Figure 2. Reliability value of each question item

The result obtained for the item with a maximum score of one (dichotomous) using the KR-20 technique was 0.59, categorized as quite reliable, as the calculated value falls within the range of 0.41-0.60. These result indicate that the items have sufficient internal consistency to measure the intended construct. Furthermore, the result obtained for the item with a maximum score of two using the Cronbach's alpha technique was 0.75, categorized reliable, as it falls within the range of 0.71-0.80. This indicates that the items provide good internal consistency and produce relatively stable measurement results. Furthermore, the result obtained for the item with a maximum score of four using the Cronbach's alpha technique was 0.82, categorized as highly reliable, as the calculated value falls within the range of 0.81-1.00. This indicates that the items consistently measure the same construct. The higher reliability coefficients obtained for polytomous items compared to dichotomous items indicate that score categories greater than 1 are able to capture variation in student responses more effectively, thereby improving measurement consistency.

These findings align with previous research reporting that assessment instruments with good internal consistency provide more stable and reliable measurement results. Boateng et al., (2018)noted that internal consistency reflects the degree of correlation between items in a scale and its total score, and an adequate level of internal consistency indicates that the items function consistently in measuring the same construct.. Similarly, Farida and Musyarofah (2021) emphasizes that higher internal consistency reliability coefficients reflect greater consistency among items in measuring the intended construct. Therefore, the high and highly reliable reliability coefficients obtained in this study indicate that the developed instrument is able to consistently assess students' problem-solving abilities, while the moderate reliability obtained for the dichotomous items suggests that further refinement of some items could improve the instrument's overall consistency.

The magnitude of the reliability value obtained can be influenced by several factors, including: 1) Number of items; 2) Clarity of instructions/guidelines; 3) Condition of the test participants; 4) Time of test administration; and 5) Suitability of the question format to the measurement objectives. The number of test items can affect reliability because too few questions are considered inadequate to fully represent students' abilities, resulting in inconsistent tests. The clarity of test instructions can also affect reliability. Unclear instructions can lead to ambiguity or confusion in students' understanding and answering the questions, leading to errors in answering them. Furthermore, student conditions, such as learning motivation, fatigue level, psychological state, or health, can affect reliability. If students are in less than optimal condition, they tend to provide

inconsistent answers. Furthermore, insufficient test time can affect students' concentration on the questions, leading to errors, which in turn lower test reliability. Furthermore, the appropriateness of the question format to the measurement objectives affects reliability. If the question format used does not align with the needs being measured, the results obtained will not consistently reflect true abilities(Muharromah et al., 2025).In this study, the relatively higher reliability coefficients obtained for the polytomous items indicate that the developed assessment rubric and item format are sufficiently consistent to minimize measurement error while capturing varying levels of student problem-solving performance.

Evaluation stage

Based on a series of planning, development, testing, and data analysis stages that have been carried out, this research has produced an instrument product that is integrated with problem-solving skills in the context of SDGs 12 which is declared **valid in content** because it has a CVR value above 0.99 and is largely **valid empirically** because the calculated r-value is greater than 0.334. The instrument is also declared **reliable** although at different levels because the calculated reliability value shows that the questions are in the range of fairly reliable to high category.

▪ CONCLUSION

This study successfully developed and validated an AKM instrument that integrates reading literacy, numeracy, and problem-solving skills within the subtopic of factors affecting reaction rates. The validity results indicate that the instrument is suitable for use, with a Content Validity Ratio (CVR) of ≥ 0.99. Furthermore, of the 30 items selected for the pilot test, 20 met the empirical validity criteria (rcount > 0.334), while 10 did not meet these criteria and therefore require revision before being applied more widely.

Reliability test results also indicate that the instrument exhibits a good level of consistency. The reliability for multiple-choice items, multiple-answer items, and short-answer items with a maximum score of one was 0.59, indicating that these items fall into the “moderately reliable” category. Meanwhile, the reliability of essay items with a maximum score of two was 0.75, which falls into the “reliable” category, while that of matching items with a maximum score of four was 0.82, indicating “highly reliable.” These results indicate that the AKM instrument developed is valid and reliable for use as a diagnostic assessment to measure students’ reading literacy, numeracy, and problem-solving skills regarding the factors affecting reaction rates.

This study contributes to the development of AKM-based chemistry assessments by presenting an instrument that integrates reading literacy, numeracy, and problem-solving skills, as well as implementation via a digital platform, all within a single assessment instrument. Thus, this instrument can serve as an alternative for teachers in conducting more contextual and competency-oriented assessments. Future research is recommended to revise the 10 items that do not yet meet empirical validity criteria and to retest the instrument on a larger and more diverse sample to improve the instrument’s quality. Additionally, this instrument can be implemented as a diagnostic or formative assessment in high school chemistry instruction to identify students literacy, numeracy, and problem-solving skills as a basis for planning follow-up learning activities.

▪ REFERENCES

Adams, W. K., & Wieman, C. E. (2010). 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

Amali, N., . N., & Firman, H. (2024). A Framework Design for Developing and Validating Virtual Test to Assess Science Process Skills in Chemistry. *KnE Social Sciences*. https://doi.org/10.18502/kss.v9i8.15557

Bankole, H. B., & Ayanwale, M. A. (2025). Digital assessments boost science attitudes and sustainable education. *Discover Sustainability*, *6*(1). https://doi.org/10.1007/s43621-025-01963-7

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, *6*. https://doi.org/10.3389/fpubh.2018.00149

Capah, B. M., Rachim, H. A., & Raharjo, S. T. (2023). *Implementasi Sdg's-12 Melalui Pengembangan Komunitas Dalam Program Csr*. https://doi.org/10.45814/share.v13i1.46502

Daeli, V., Yakin Niat Telaumbanua, Sadiana Lase, & Ratna Natalia Mendrofa. (2025). Analisis Kemampuan Numerasi Siswa Dalam Menyelesaikan Soal Asesmen Kompetensi Minimum di SMA Negeri 1 Gunungsitoli. *Emasains : Jurnal Edukasi Matematika Dan Sains*, *14*(1), 59–77. https://doi.org/10.59672/emasains.v14i1.4553

Fahmy, A., Apriliyani, D., Janah, R., & Islam Negeri Abdurrahman Wahid Pekalongan, U. K. (2025). *Literasi Numerasi Siswa SMP dalam Menyelesaikan Soal Asesmen Kompetensi Minimum*. *5*(2). http://e-journal.uingusdur.ac.id/index.php/circle

Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, *I*. https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB

Gliem, J. A., & Gliem, R. R. (2003). *Midwest Research to Practice Conference in Adult, Continuing, and Community Education*.

Halim, A. (2022). Signifikansi Dan Implementasi Berpikir Kritis Dalam Proyeksi Dunia Pendidikan Abad 21 Pada Tingkat Sekolah Dasar. *Jurnal Indonesia Sosial Teknologi*.

Handayani. (2025). Implementasi Pembelajaran Mendalam (Deep Learning) Dalam Meningkatkan Efektivitas Kurikulum Berdampak Di Sekolah. *Jurnal Edu Research*, *6*(2).

Hidayatulloh, R., & Utiya Azizah, dan. (2020). Analisis Keterampilan Pemecahan Masalah Siswa Sma Pada Topik Laju Reaksi. *Jurnal Penelitian Pendidikan Sains)*, *10*(01). https://journal.unesa.ac.id/index.php/jpps

Inayah, A., Nagari, Setiawan, & Anisah. (2022). Kemampuan Literasi Numerasi untuk Mengembangkan Pemahaman Konsep dalam Pembelajaran Matematika. *Jurnal Ilmiah Mahasiswa Pendidikan Matematika*.

Irmayta, E., Rudibyani, R. B., & Efkar, T. (2017). Pengembangan Instrumen Asesmen Pengetahuan pada Materi Asam Basa Arrhenius. *Jurnal Pendidikan Dan Pembelajaran Kimia*, *7*.

Islamiyah, & Yuliani. (2024). Validitas Dan Reliabilitas Instrumen Tes Miskonsepsi Peserta Didik Menggunakan Metode Four-Tier Diagnostic Test Pada Konsep Respirasi. *BioEdu*. https://ejournal.unesa.ac.id/index.php/bioedu

Kementerian Pendidikan Dasar dan Menengah. (2025). *Rapor Pendidikan Indonesia 2025*.

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

Lestari. (2021). Identifikasi Miskonsepsi Siswa pada Materi Laju Reaksi dan Perbaikannya Menggunakan Model Pembelajaran Learning Cycle 5E dengan Strategi Konflik Kognitif. *Jurnal Pendidikan*, *6*. http://journal.um.ac.id/index.php/jptpp/

Lubay, O., Fadiawati, N., & Rosilawati, I. (2024). *Enhancing Students' Analyzing Skills on Reaction Rate Topics Using the Guided Inquiry Learning Model*. *13*(1), 162–169. https://doi.org/10.23960/jppk.v13.i1.2024.16

Muharromah, F., Fani, F., & Ridho, A. (2025). *Reliabilitas Alat Ukur: Jenis-Jenis, Cara Pengukuran, Dan Faktor- Faktor Yang Mempengaruhinya Dalam Pembelajaran*.

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

Purniasari, L., Masykuri, M., Sri, D., & Ariani, R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia Sma N 1 Kutowinangun Tahun Pelajaran 2019/2020 Menggunakan Model Iteman Dan Rasch. *Jurnal Pendidikan Kimia*, *10*. https://doi.org/10.20961/jpkim.v10i2.48244

Pusmendik. (2023). *Framework Asesmen Kompetensi Minimum (AKM)*.

Rismana, N., & Hernawati, D. S. (2025). Pengembangan Kurikulum di Indonesia Dalam Menghadapi Tuntutan Abad Ke-21. *Jurnal Pendidikan Islam*, *12*(1), 2025.

Riyani, R., & Maizora, S. (2017). Uji Validitas Pengembangan Tes Untuk Mengukur Kemampuan Pemahaman Relasional Pada Materi Persamaan Kuadrat Siswa Kelas VIII Smp. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, *1*(1), 2581–253.

Rokhim, Tyas, Rahayu, & Habiddin. (2022). Perspektif Siswa Dan Guru Dalam Pelaksanaan Akm (Asesmen Kompetensi Minimum) Pada Mata Pelajaran Kimia. *JAMP: Jurnal Adminitrasi Dan Manajemen Pendidikan*, *5*. http://journal2.um.ac.id/index.php/jamp/

Rosmiati. (2022). Pembelajaran Kimia Yang Menyenangkan Di Madrasah. *Uniqbu Journal of Exact Sciences*.

Rosnaeni, R. (2021). Karakteristik dan Asesmen Pembelajaran Abad 21. *Jurnal Basicedu*, *5*(5), 4341–4350. https://doi.org/10.31004/basicedu.v5i5.1548

Saputri, D., Hidayati, N., Fauziah, N., & Artikel Abstrak, I. (2023). Lembar Validasi: Instrumen yang Digunakan Untuk Menilai Produk yang Dikembangkan Pada Penelitian Pengembangan Bidang Pendidikan. *Biology And Education Journal*, *3*(2), 133–151.

Son. (2019). *Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal*.

Subhaktiyasa. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka. *Journal of Education Research*.

Supiati, & Mujib, A. (2025). *Identifikasi Miskonsepsi Siswa Menggunakan Three Tier Test Pada Materi Teorema Pythagoras*.

Suwahono, Sa'adah, & Yulianingsih. (2025). Diagnostic Instruments For Numeracy Skills In Chemistry Learning: A Development Study Oriondo AntoniO. *EDUSAINS*, *17*. https://doi.org/10.15408/es.v17i2.49862

Zahroh. (2019). Pengembangan Instrumen Soal Berorientasi Praktek Keterampilan Pemecahan Masalah Materi Laju Reaksi. *JRPK: Jurnal Riset Pendidikan Kimia*, *13*(1), 78. https://doi.org/10.21009/jrpk.131.10