Tersedia online di EDUSAINS
Website: http://journal.uinjkt.ac.id/index.php/edusains

EDUSAINS, 17 (2), 2025, 146-154

Research Artikel

DIAGNOSTIC INSTRUMENTS FOR NUMERACY SKILLS IN CHEMISTRY LEARNING: A DEVELOPMENT STUDY ORIONDO ANTONIO

Suwahono[1*], Fina Sa'adah[2], Yulianingsih[3]
[1,2,]UIN Walisongo, Semarang, Indonesia
[3]Uni-Freiburg-de, Germany
suwahono@walisongo.ac.id [1*]

Abstract

This study aims to develop a valid and reliable diagnostic instrument to measure students' numeracy skills in the context of chemistry learning. The instrument development followed a systematic procedure based on the Oriondo & Antonio model, which includes the stages of planning, construction, validation, and revision. The instrument was designed based on numeracy indicators integrated with chemistry content, particularly on topics requiring quantitative understanding, such as reaction rates. Content validity was reviewed by chemistry and education experts, while empirical reliability testing was conducted using Rasch modelling. The analysis results showed that the instrument had an Aiken validity of 0.80 (valid category), item reliability of 0.88 (very good category), person reliability of 0.80 (good category), and Cronbach's alpha value of 0.75 (good category). These findings indicate that the developed instrument meets the criteria for validity and reliability and is capable of specifically identifying students' misconceptions and numeracy gaps. Therefore, this instrument has the potential to be used as an initial diagnostic tool in designing remedial or enrichment learning strategies in chemistry classes.

Keywords*: Numeracy skills, diagnostic instrument, chemistry learning, validity, reliability*

Permalink/**DOI:** https://doi.org/10.15408/es.v17i2.49862

How To Cite: Suwahono, S., Sa'adah, F., Yulianingsih, Y. (2025). Diagnostic Instruments for Numeracy Skill as in Chemistry Learning: a Development Study Oriondo Antonio. *EDUSAINS*, 17 (2): 146-154.

*Corresponding author

Received: 21 December 2025; Revised: 22-31 December 2025; Accepted: 31 December 2025

INTRODUCTION

The 21st century demands that individuals possess strong essential competencies to actively and productively participate in an increasingly complex, dynamic, and data-driven society. One of the primary competencies is numeracy, which involves not only basic arithmetic skills but also the ability to think logically, analyze data, and solve contextual problems using quantitative information (Safitri et al., 2024). Numeracy is defined as the ability to apply number concepts and mathematical operations, as well as the ability to read, interpret, and use graphs, tables, and numerical data to make informed decisions (Ananiadou et al., 2009; Mullis et al., 2021)

The importance of numeracy is emphasized in national education policy through the implementation of the Minimum Competency Assessment (AKM), which positions numeracy, alongside reading literacy, as the foundation for lifelong learning (Education Assessment Centre, 2023). This skill is not only required in mathematics but also plays a crucial role in chemistry education, particularly in topics involving quantitative analysis and numerical data processing (Morel & Morgan, 1972). Topics such as reaction rates, stoichiometry, solution concentration, and gas laws require students to understand the mathematical relationships between variables and apply them in experimental contexts and chemical problem-solving (Bravenec & Ward, 2023; Moruk & Sulisworo, 2024).

Studies show that students often struggle to transfer their numerical skills to a chemistry context, leading to misconceptions and errors in interpreting scientific data (Ariani et al., 2024). Strong numeracy skills are a prerequisite for students to understand and master chemistry concepts in depth (Aldossary et al., 2024). However, in reality, many students still struggle to integrate their understanding of chemical concepts with the necessary numerical skills, which impacts their learning outcomes and conceptual understanding (Nguyen et al., 2021). This condition highlights the need to strengthen the integration of numeracy in chemistry learning through an approach that is not only outcome-oriented but also capable of detecting students' weaknesses and misconceptions early on (Suparman et al., 2024). Therefore, a diagnostic approach is an appropriate alternative for specifically identifying gaps in students' numerical understanding, enabling teachers to design more adaptive learning strategies, both for remediation and enrichment.

In response to the need for integrating numeracy into chemistry education, this study focuses on developing a diagnostic instrument for numeracy skills specifically designed for reaction rate material. The selection of this material is based on its high level of complexity, as it involves abstract concepts, mathematical calculations, and the ability to interpret experimental data such as concentration, time, and reaction rate (D'Alessio et al., 2020). The main advantage of this instrument lies in its open-ended question format, which allows for in-depth exploration of students' error patterns and difficulties in applying numeracy concepts in a chemistry context (Üce & Ceyhan, 2019). This diagnostic approach differs from conventional formative or summative tests, as it is designed to uncover detailed information about students' understanding gaps, enabling teachers to design more targeted and effective learning interventions (Sayre et al., 2025)

Although the importance of numeracy has been widely recognized, and various general numeracy measurement instruments are available, this study identifies a significant gap, namely the lack of diagnostic instruments specifically designed to measure numeracy skills in the context of chemistry learning in Indonesia. Existing instruments tend to be general or focus solely on final outcomes, without providing an in-depth understanding of students' thinking processes or the conceptual error patterns they experience. As a result, teachers struggle to obtain detailed information that can be used to design learning tailored to each student's needs.

The novelty of this research lies in the development of a diagnostic instrument for numeracy skills in reaction rate material using the Oriondo & Antonio development model. This model provides a systematic framework that includes instrument design, expert validation, field testing, and empirical data analysis using the Rasch Model approach. The use of the Rasch Model enables more accurate evaluation of item characteristics, such as difficulty level and discriminative power, as well as students' abilities through reliability estimation and response patterns (Wright, 1977). Thus, the resulting instrument not only meets statistical validity and reliability criteria but also provides in-depth diagnostic insights into students' numeracy profiles in the context of chemistry learning (Easa & Blonder, 2022). Through the use of this instrument, teachers are expected to diagnose more accurately the location of students' numeracy difficulties on the topic of reaction rates, so that remedial and enrichment learning strategies can be designed more effectively and tailored to their needs.

METHOD

This study adopted a Research and Development (RnD) approach using the Oriondo Antonio development model, which was chosen for its systematic and comprehensive nature in instrument development (Oriondo & Antonio, 1984). The development of diagnostic instruments for numeracy skills in this study was carried out in three main stages, namely: (1) test design, (2) product testing, and (3) final instrument assembly.

Test Design, The initial stage began with determining the purpose of the instrument, which was to diagnose students' numeracy skills in understanding reaction rate material. This material was chosen because of its high complexity, encompassing conceptual, quantitative, and interpretative aspects, particularly in analysing the relationship between concentration, time, and reaction rate. Next, an instrument grid was developed, containing important components such as basic competencies, question indicators, and numeracy ability indicators. The numeracy indicators referenced include the ability to use numbers and symbols, analyse quantitative information, and interpret analysis results in a scientific context. Based on this framework, 25 open-ended questions were developed to assess students' numeracy skills in contextual chemistry situations. After the questions were developed, content validation was conducted by two chemistry education lecturers and three experienced high school chemistry teachers specialising in instrument development. Validation was conducted on the content, construct, and language aspects using a validation sheet with a specific rating scale. The validation results were analysed using Aiken's V formula to determine the level of agreement among experts on the suitability of each question item (Merino-Soto, 2023). Items that received low scores or received feedback for improvement from the validators were then revised. At this stage, scoring guidelines (assessment rubrics) were also developed for each item to ensure objectivity and consistency in the assessment process.

Product Testing, The revised instrument is then pilot-tested on 30 Grade XII students at one of the state high schools in Semarang. The purpose of this pilot test was to obtain empirical data for use in analysing question quality, including difficulty level, discriminative power, and reliability. The analysis was conducted using Rasch modelling through the Winsteps application. Additionally, a readability test was conducted with students to evaluate how well the language, terms, and instructions in the questions could be understood by the students.

Final Instrument Development, After the pilot test results were analysed and revisions were made based on empirical findings, the final version of the instrument was developed. This instrument is now ready to be used to identify students' numeracy skills in reaction rate material in a more in-depth and diagnostic manner. The instrument's strength lies in its ability to reveal students' thinking errors and specific weaknesses in integrating chemistry concepts with numeracy skills.

The data from the instrument trial was analysed quantitatively using Rasch Modelling with the assistance of Winsteps software. Rasch Model analysis allows for a careful evaluation of the quality of the instrument, including: (1) Item Validity (Item Fit): Determining whether each item behaves in accordance with the Rasch model, indicating whether the item measures the same construct as other items. Items that fit are considered statistically valid. (2) Item and Person Reliability: Measuring the consistency of the instrument's measurement (item reliability) and how well the instrument can distinguish between students' ability levels (person reliability). Cronbach's Alpha values are also calculated as an indicator of the instrument's internal consistency. (3) Item Difficulty: Determining how difficult each item is for students, estimated based on the proportion of students who answered correctly. Items are categorised as very difficult, difficult, moderate, easy, and very easy. (4) Item Discrimination: Measuring the ability of each item to distinguish between high- and low-ability students.

RESULTS AND DISCUSSION

The development process of the diagnostic instrument for numeracy skills in reaction rate material followed the Research and Development (RnD) model by Oriondo Antonio. The initial stage, namely test design, began with determining the instrument's objectives: to measure students' numeracy skills, including indicators of number/symbol usage, quantitative information analysis, and interpretation of analysis results, particularly in the context of reaction rate material (covering reaction rate concepts, influencing factors, and reaction order calculations). The selection of the Oriondo Antonio model provides a systematic framework, ensuring that the instrument is developed through measurable and planned stages, from needs identification to product validation (Abdurrahman & Mahmudah, 2023).

Next, a question matrix was developed to serve as a blueprint for writing question items. This matrix maps Core Competencies into specific question indicators and links them to numeracy ability indicators, which are essential to ensure that the scope of the material and numeracy aspects are covered. From this framework, 25 initial essay questions were developed. Essay questions were chosen because they allow for the exploration of students' thinking processes and reasoning in solving numerical problems in chemistry, unlike multiple-choice questions, which tend to only measure the final answer (Pradana et al., 2023).

The next stage was expert validation to assess the quality of the instrument in terms of content, construct, and language. Validation was conducted by two chemistry education lecturers and three experienced high school teachers, who provided crucial feedback for improvements. This expert validation process is important to ensure the instrument has high content and construct validity before being empirically tested (Aiken, 1980). The expert validation results in Table 1 show that most questions met the eligibility criteria, with minor revisions made based on validator suggestions. The Aiken's V validity test result of 0.88 > 0.80 falls within the valid category (Merino-Soto, 2023).

Table 1. Expert validation results

Aspects Evaluated	Average	Score
Instructions	90	Very good
Numbering system	90	Very good
Font type and size	95	Very good
Tables/graphs/images	83	Very good
Use of sentences	85	Very good
Easy understanding questions on numerical ability in reaction rate material	88	Very good
Likelihood of questions being solvable	85	Very good
Average	88	Very good

After revision based on expert validation, the instrument was tested on 63 grade XII students majoring in Mathematics and Natural Sciences at a public high school in Semarang. The test results were then analysed using Rasch modelling. Initial identification took the form of a dimensionality test of the developed items. The results of the unidimensionality of the developed test items were good (Table 2). These results indicate that the developed items have the ability to objectively estimate individual ability and item characteristics, independent of the sample used, and provide rich diagnostic information about item behaviour (Huang et al., 2023). This analysis allows the determination of which items are "fit" or consistent with the model, making them suitable for use as measurement instruments.

Tabel 2. Results of dimensionality identification in Eigenvalue units

No	Standardized Residual Variance	Value		Category
		Empirical	Modelled	
1	Raw variance explained by measures	67.3	64.4	Special
2	Raw variance explained by persons	31.0	29.7	Good
3	Raw variance explained by items	36.3	34.7	Good
4	Raw unexplained variance (total)	32.7	35.6	Good

Based on the Rasch Model analysis Table 2, Standardized Residual Variance shows that the Rasch model is able to explain most of the data variation in the conceptual test instrument with polytomous scoring. The raw variance explained by measures reached 67.3% in the empirical data and 64.4% in the model data, indicating an excellent level of model fit. This value exceeds the minimum recommended threshold to meet the one-dimensionality assumption, and the 25 essay questions tested, 15 were deemed suitable and appropriate for use. The items that did not fit indicated deviations from the model, such as questions that were too difficult or too easy for most respondents, or had low discriminative power. This is consistent with the person fit analysis conducted on 134 response sheets, where 88.05% or 118 responses were considered normal and aligned with the model, while 11.95% or 16 responses showed deviations from the model. These deviations indicate that not all items or responses are reliable or in accordance with the expectations of the Rasch model, which emphasizes the need for stringent selection of items used in the assessment. This finding aligns with the work of Fauzi et al. (2022), which demonstrated that not all designed items will consistently align with psychometric models and require strict selection to ensure the validity of assessment instruments.

This is in line with other studies showing that not all items designed will always fit the psychometric model and require strict selection (Fauzi et al., 2022). Analysis of the response patterns of assessors to the level of difficulty of the items revealed a varied distribution: 3 items in the very difficult category, 7 items in the difficult category, 4 items in the easy category, and 4 items in the very easy category. This distribution indicates that the instrument has an adequate range of difficulty, enabling it to measure students' abilities at various levels. Questions that are too easy or too difficult do not provide enough information about students' abilities, so variation in difficulty levels is a desirable characteristic for diagnostic instruments. A good question is one that can showcase students' abilities and skills at a certain level, with the ability to distinguish between students who have a better understanding and those who need improvement. Therefore, variation in the difficulty level of questions is crucial to ensure that assessment instruments can provide an accurate picture of students' ability levels (Ramadhan et al., 2023). The response patterns based on the difficulty levels of the developed questions fall into the category of Stochastically Modeled Diagnostics (Linacre & Wright, 1994).

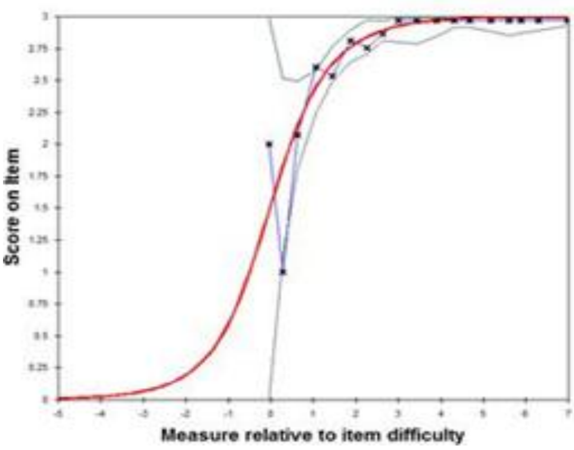

Figure 1. response patterns of assesses to the difficulty level of items for the developed items

Figure 1, The sigmoid curve shown in the graph indicates that the item has good psychometric characteristics. At low ability levels ($\theta < -2$), the probability of students obtaining a high score is very small, indicating that students with low numeracy abilities tend not to be able to solve the problem correctly. This is in line with the instrument's purpose as a diagnostic tool capable of identifying initial difficulties and misconceptions about numeracy in the context of reaction rate material. The reliability aspects of the instrument were carefully evaluated (Table 3). The item reliability of the instrument showed a value of 0.88, which falls into the good category. This figure indicates that the items in this instrument consistently measure numeracy ability (Hakkarainen et al., 2023). However, the person reliability showed a value of 0.75, which also falls into the good category. Nevertheless, the lower person reliability value (compared to item reliability) may indicate that the instrument is less able to consistently distinguish between the ability levels of individual students in the test sample (Chin & Chew, 2023). This may be due to a relatively homogeneous sample or the presence of unpredictable response variations (Patac et al., 2021). Nevertheless, Cronbach's Alpha value of 0.75, which is categorized as adequate, indicates that the instrument has acceptable internal consistency (Ahmad et al., 2024). This internal consistency is crucial to ensure that the instrument is reliable in measurement (Pentapati et al., 2025).

Table 3. Reliability Test Results

Reliability	**Value**	**Category**
Person Reliability	0.80	Good
Item Reliability	0.88	Good
Cronbach's alpha	0.75	Good

Furthermore, item discrimination shows that there are 4 groups of items, while respondent discrimination shows 2 groups of respondents. Good item discrimination is important to ensure that the items can distinguish between high-ability and low-ability students. The more groups of items discriminate, the better the instrument is at identifying differences in ability. Item discrimination refers to a question's ability to differentiate between students with high and low abilities (Surhasimi & Arikunto, 2016). Item fit testing provides information that the response patterns of the 3 categories of assessed ability levels are in 3 patterns: low, medium, and high.

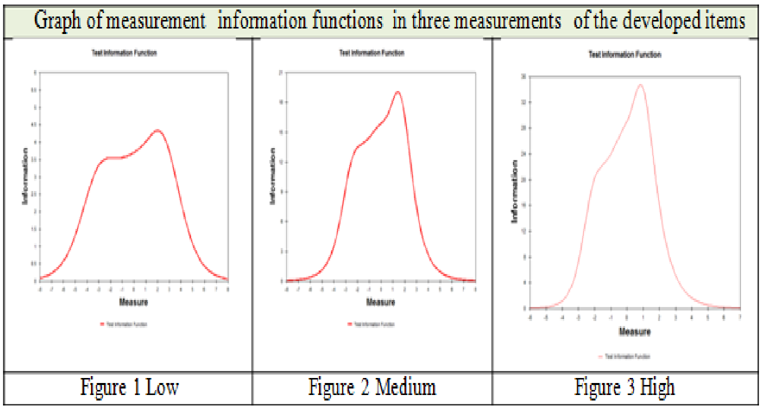

Figure 1 Low

Figure 2 Medium

Figure 3 High

Figure 2. Measurement information graph models

Figure 2 illustrate the response patterns of students to questions based on difficulty levels analyzed using the Rasch Model. For questions with low difficulty levels (b low), participants with low ability (θ low) still have a high chance of answering correctly. Overall, these three Measurement Information Function graphs confirm that the developed diagnostic numeracy instrument has comprehensive measurement coverage across various student ability levels. The instrument performs best at medium-sized ability levels but remains informative at low and high levels. This finding is consistent with previous reliability, discrimination, and item fit results and strengthens the conclusion that the instrument is suitable for diagnostically mapping students' numeracy ability profiles in chemistry learning, particularly in reaction rate material. Participants with medium to high ability almost always answer correctly, as the questions are below their ability level. This is reflected in the Item Characteristic Curve (ICC), which tends to be steep on the left side, indicating a high overall success rate (Andersen, 1973). For items with moderate difficulty, the response pattern shows that participants with low ability tend to answer incorrectly, participants with moderate ability have approximately a 50% chance of answering correctly, and participants with high ability generally answer correctly. Items in this category are highly effective in distinguishing participants' abilities, as they are positioned at the midpoint of the ability scale (θ = b), resulting in the ICC curve intersecting the 0.5 probability point.

Meanwhile, for high-difficulty questions, only high-ability participants (high θ) have a high probability of answering correctly. Low- and medium-ability participants show low probabilities of correct answers, reflecting the high selectivity of the questions toward high-ability participants (Andersen, 1973). Overall, this instrument demonstrates strong diagnostic capabilities in mapping students' ability profiles probabilistically. Through these response patterns, teachers can identify areas of basic mastery, identify initial misconceptions, and obtain detailed information about numeracy concepts that students have not yet mastered in the context of chemistry learning.

CONCLUSION

This study successfully developed a valid and reliable diagnostic instrument to measure students' numeracy skills in the context of chemistry learning. The developed instrument can identify variations in students' numeracy skills, detecting specific misconceptions, and demonstrating high discriminating power. The theoretical contribution of this study lies in the integration of numeracy concepts and chemistry content in the form of a diagnostic assessment based on the Rasch Model. Meanwhile, its practical contribution provides concrete solutions for chemistry teachers in designing adaptive learning strategies, both for remedial and enrichment purposes. Therefore, this instrument has strong potential to be used as an initial diagnostic assessment tool in the learning process. Further development is recommended to expand the scope of the material and increase sample heterogeneity to strengthen the instrument's ability to differentiate students' ability levels more comprehensively.

ACKNOWLEDGMENT

The author sincerely thanks the Rector of UIN Walisongo Semarang for the approval and financial support provided for the implementation of this programme development research. This support is highly meaningful and has significantly contributed to the success of the research.

REFERENCES

Abdurrahman, I. S., & Mahmudah, F. N. (2023). Development of a Digital-Preneurship Measurement Instrument: Alignment Approach Through Project-Based Learning. *International Journal of Educational Methodology*, *9*(1), 283–295. https://doi.org/10.12973/IJEM.9.1.283

Achmad Rante Suparman, Eli Rohaeti, & Sri Wening. (2024). Student Misconception In Chemistry: A Systematic Literature Review. *Pegem Journal of Education and Instruction*, *14*(2). https://doi.org/10.47750/pegegog.14.02.28

Ahmad, Alias, Hamat, & Mohamed. (2024). RELIABILITY ANALYSIS: APPLICATION OF CRONBACH'S ALPHA IN RESEARCH INSTRUMENTS. *Pioneering the Future: Delving Into E-Learning's Landscape*, 114–119. https://appspenang.uitm.edu.my/sigcs/2024-2/Articles/20244_ReliabilityAnalysis-ApplicationOfCronbachsAlphaInResearchInstruments.pdf

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Aldossary, A., Campos-Gonzalez-Angulo, J. A., Pablo-García, S., Leong, S. X., Rajaonson, E. M., Thiede, L., Tom, G., Wang, A., Avagliano, D., & Aspuru-Guzik, A. (2024). In Silico Chemical Experiments in the Age of AI: From Quantum Chemistry to Machine Learning and Back. In *Advanced Materials* (Vol. 36, Issue 30). John Wiley and Sons Inc. https://doi.org/10.1002/adma.202402369

Ananiadou, K., Claro, M., & Magdalean Claro, oecdorg. (2009). 21st Century Skills and Competences for New Millennium Learners in OECD Countries. *OECD Education*

Working Papers, *2009*(41), 33. https://doi.org/10.1787/218525261154

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, *38*(1). https://doi.org/10.1007/BF02291180

Ariani, Y., Suparman, Helsa, Y., Zainil, M., & Rahmatina. (2024). ICT-Based or-Assisted Mathematics Learning and Numerical Literacy: A Systematic Review and Meta-Analysis. *International Journal of Information and Education Technology*, *14*(3). https://doi.org/10.18178/ijiet.2024.14.3.2060

Bravenec, A. D., & Ward, K. D. (2023). Interactive Python Notebooks for Physical Chemistry. *Journal of Chemical Education*, *100*(2). https://doi.org/10.1021/acs.jchemed.2c00665

Chin, H., & Chew, C. M. (2023). Cognitive diagnostic assessment with ordered multiple-choice items for word problems involving 'Time.' *Current Psychology*, *42*(20). https://doi.org/10.1007/s12144-022-02965-8

D'Alessio, G., Parente, A., Stagni, A., & Cuoci, A. (2020). Adaptive chemistry via pre-partitioning of composition space and mechanism reduction. *Combustion and Flame*, *211*. https://doi.org/10.1016/j.combustflame.2019.09.010

Easa, E., & Blonder, R. (2022). Development and validation of customized pedagogical kits for high-school chemistry teaching and learning: the redox reaction example. *Chemistry Teacher International*, *4*(1). https://doi.org/10.1515/cti-2021-0022

Education Assessment Centre. (2023). *National AKM Report*. Ministry of Education, Culture, Research, and Technology. https://anbk.kemdikbud.go.id/anbk2023/

Fauzi, A. A., Susongko, P., & Hayati, M. N. (2022). Tes Kemampuan Berpikir Kritis pada Pembelajaran IPA di SMP Berbasis Model Rasch. *PSEJ (Pancasakti Science Education Journal)*, *7*(1). https://doi.org/10.24905/psej.v7i1.146

Hakkarainen, A., Cordier, R., Parsons, L., Yoon, S., Laine, A., Aunio, P., & Speyer, R. (2023). A systematic review of functional numeracy measures for 9–12 -year-olds: Validity and reliability evidence. *International Journal of Educational Research*, *119*. https://doi.org/10.1016/j.ijer.2023.102172

Huang, L., Shu, X., Ge, N., Gao, L., Xu, P., Zhang, Y., Chen, Y., Yue, J., & Wu, C. (2023). The accuracy of screening instruments for sarcopenia: a diagnostic systematic review and meta-analysis. In *Age and Ageing* (Vol. 52, Issue 8). https://doi.org/10.1093/ageing/afad152

Linacre, J., & Wright, B. D. (1994). Dichotomous Mean Square Chi-square fit statistics. *Rasch Measurement Transactions1*, *8*(2).

Merino-Soto, C. (2023). Aiken's V Coefficient: Differences in Content Validity Judgments. *MHSalud*, *20*(1). https://doi.org/10.15359/mhs.20-1.3

Morel, F., & Morgan, J. (1972). A Numerical Method for Computing Equilibria in Aqueous Chemical Systems. *Environmental Science and Technology*, *6*(1). https://doi.org/10.1021/es60060a006

Moruk, S., & Sulisworo, D. (2024). Literature Review on Longitudinal Study of Improving Numerical Literacy at Elementary Education. *Buletin Edukasi Indonesia*, *3*(03). https://doi.org/10.56741/bei.v3i03.757

Mullis, Martin, & Davier, V. (2021). TIMSS 2023 Assessment Framework. In *TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.*

Nguyen, H. T., Domingo, P., Vervisch, L., & Nguyen, P. D. (2021). Machine learning for integrating combustion chemistry in numerical simulations. *Energy and AI*, *5*. https://doi.org/10.1016/j.egyai.2021.100082

Oriondo, L. Loyola., & Antonio, E. M. D.-. (1984). *Evaluating educational outcomes : tests, measurement and evaluation*. Rex Book Store.

Patac, L. P., Adriano, Jr. P., & Bactil, C. M. (2021). Factor analytic method in developing scoring rubric for word problems. *Asia Research Network Journal of Education*, *1*(3).

Pentapati, K. C., Chenna, D., Kumar, V. S., & Kumar, N. (2025). Reliability generalization meta-analysis of Cronbach's alpha of the oral impacts on daily performance (OIDP) questionnaire. In *BMC Oral Health* (Vol. 25, Issue 1). https://doi.org/10.1186/s12903-025-05496-3

Pradana, P. W., Febriani, F., Ibnusaputra, M., & Jumadi, J. (2023). Development of Physics Test Instrument to Measure Verbal Representation of High School Student on Optical Instrument Topic. *Jurnal Penelitian Pendidikan IPA*, *9*(10). https://doi.org/10.29303/jppipa.v9i10.3775

Ramadhan, W., Malahati, F., Romadhon, K., & Ramadhan, S. (2023). Analisis Butir Soal Tipe Multiple Choice Questions pada Penilaian Harian Sekolah Dasar. *Tarbiyah Wa Ta’lim: Jurnal Penelitian Pendidikan Dan Pembelajaran*, *10*(2). https://doi.org/10.21093/twt.v10i2.6155

Rini Rahma Safitri, Gita Asyari, Dara Avira, & Abdul Fattah Nasution. (2024). Rekonstruksi Minat Belajar Peserta Didik Abad 21 Melalui Model Sistem Dinamis. *Student Scientific Creativity Journal*, *3*(1), 133–143. https://doi.org/10.55606/sscj-amik.v3i1.4785

Sayre, J., Nabua, E., Salic-Hairulla, M., Alcopra, A., & Fernandez, M. J. (2025). Assessing General Chemistry Learning Gaps: A Needs Assessment of Competency Mastery among Grade 11 Learners. *International Journal of Research and Innovation in Social Science*, *IX*(IV). https://doi.org/10.47772/ijriss.2025.90400472

Surhasimi, & Arikunto. (2016). Prosedur Penelitian : Suatu Pendekatan Praktik. *Rineka Cipta*, *2006*(2006).

Üce, M., & Ceyhan, İ. (2019). Misconception in Chemistry Education and Practices to Eliminate Them: Literature Analysis. *Journal of Education and Training Studies*, *7*(3). https://doi.org/10.11114/jets.v7i3.3990

Wright, B. D. (1977). SOLVING MEASUREMENT PROBLEMS WITH THE RASCH MODEL. *Journal of Educational Measurement*, *14*(2). https://doi.org/10.1111/j.1745-3984.1977.tb00031.x