JPPIPA 12 (7) (2026)

Jurnal Penelitian Pendidikan IPA

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

Developing a Chemical Literacy Test Instrument on Buffer Solutions to Support Sustainable Development Goal (SDG) 4: Quality Education

Siska Handa Yani[1], Yerimadesi[1*], Fauzana Gazali[1], Riga[1]

[1] Department of Chemistry, Faculty of Mathematics and Sciences, Universitas Negeri Padang, Padang, Indonesia.

Received: March 12, 2026
Revised: June 28, 2026
Accepted: July 25, 2026
Published: July 31, 2026

Corresponding Author:
Yerimadesi
yeri@fmipa.unp.ac.id

DOI: 10.29303/jppipa.v12i7.15677

Abstract: Chemical literacy is a crucial 21st-century competence; however, its assessment instruments, particularly for buffer solution topics, remain limited. This Research and Development (R&D) study utilizing the Rasch model was conducted to develop a chemical literacy test instrument in support of SDGs 4: Quality Education. The test was developed from eight initial items and piloted on 72 senior high school students. Data analysis techniques included the Content Validity Index (CVI) and Ministeps software analysis. The content validity results yielded a CVI value of 0.96 (valid), and empirical analysis demonstrated that all eight items satisfied the item fit criteria. The instrument demonstrated a Cronbach's alpha of 0.83 (good), person reliability of 0.78 (fair), and item reliability of 0.90 (good). These data indicate an excellent person-item interaction, moderately stable student responses, and high instrument reliability. The item separation value (2.99) classified the item difficulty into four strata, while the person separation value (1.91) grouped student abilities into three strata. Although this instrument provides comprehensive psychometric quality to support SDGs 4 targets and assists teachers in measuring students' chemical literacy, future research is recommended to modify the instrument by incorporating more complex and challenging items to capture a broader spectrum of student abilities.

Keywords: 21st-century learning; Buffer solutions; Chemical literacy; Rasch model; SDG 4

Introduction

21st-century education demands a strategic priority on the mastery of chemical literacy as a foundational pillar for students to deeply comprehend the phenomena of matter and energy (Barus, 2024). This imperative aligns with the core competencies emphasized by the Partnership for 21st Century Skills (Maisyaroh, 2023). Chemical literacy proficiency is paramount not only for conceptual understanding but also for cultivating students' scientific character (Pertiwi et al., 2018). Optimizing the achievement of Sustainable Development Goal 4 (SDG 4) requires mapping a clear relational pattern between the learning process and quality assessment systems (Ferguson & Roofe, 2020). In this context, chemical literacy test instruments serve as highly effective tools to measure cognitive abilities and the attainment level of students' learning objectives (Istyadji & Sauqina, 2023), applied through both formative and summative assessments (Atta et al., 2020).

For almost the past 20 years, Indonesia's science literacy performance has shown no significant improvement (OECD, 2023). In practice, literacy implementation across the nation has yet to function optimally (Widya et al., 2022). This low chemical literacy is evident across various topics, such as reaction rates (Diyanah & Yusmaita, 2022), buffer solutions (Fitriyani et al., 2022), green chemistry (Putra, 2022), and chemical equilibrium (Mayasri et al., 2023). Students frequently recognize chemical terminology but lack a profound

How to Cite:
Yani, iska H., Yerimadesi, Gazali, F., & Riga. (2026). Developing a Chemical Literacy Test Instrument on Buffer Solutions to Support Sustainable Development Goal (SDG) 4: Quality Education. *Jurnal Penelitian Pendidikan IPA*, *12*(7), 502–512. Retrieved from https://jppipa.unram.ac.id/index.php/jppipa/article/view/15677

conceptual understanding (Zebua & Sinaga, 2021). Furthermore, while they can successfully solve mathematical calculations, they often fail to provide correct qualitative or conceptual explanations for the underlying chemical phenomena (Mutiah & Siahaan, 2022). This issue is further exacerbated by imbalances in literacy dimensions within chemistry textbooks (Mustika et al., 2020), and a lack of proficiency among teachers regarding chemical literacy instruments, which leads to a scarcity of adequate testing materials (Yusmar & Fadilah, 2023; Fuadi et al., 2020). Consequently, students remain unaccustomed to chemical literacy assessment formats (Rahmasari et al., 2022) and struggle to identify relevant chemical information (Khaeruman et al., 2025), even though high chemical literacy directly correlates with the mastery of higher-order thinking skills (HOTS) (Purwanti et al., 2022). Ultimately, educational systems must equip students with the skills necessary to solve global problems through chemical literacy to realize sustainable educational quality (Sjöström et al., 2024; Sumarni et al., 2023).

Within this context, one of the conceptual topics that is highly crucial yet challenging for students' chemical literacy is buffer solutions, which are characterized by their abstract and complex nature, concurrently involving three levels of chemical representation. This topic is heavily dominated by potential misconceptions, particularly regarding the reasoning behind the mechanisms of buffer solutions in maintaining pH against the addition of small amounts of acids, bases, or dilution (Nurhidayatulah & Prodjosantoso, 2018). Nevertheless, mastering this material is a vital prerequisite for understanding various real-world phenomena and global issues (Sariati et al., 2020). Equipping students with the skills to solve global problems needs to be integrated into phenomenon-based contextual test instruments (Muntholib et al., 2020). Ideally, the characteristics of a chemical literacy test instrument must be grounded in real-world contexts and not merely evaluate the memorization of mathematical formulas (Melati et al., 2024; Siska & Yerimadesi, 2023).

Fulfilling the SDG 4 targets demands concrete strategies that focus not only on theory but also on providing learning measurement tools that are adaptive to global challenges (Ferguson & Roofe, 2020). Processing chemical data is key to cultivating sharpness of understanding (Yerimadesi et al., 2019; Yerimadesi et al., 2023), enabling students to comprehend the material within its context through numeracy literacy and HOTS (Sutra & Tantra, 2020). The achievement of sustainable development goals obligates the educational system to equip students with genuine competencies to solve environmental problems (Ghanem, 2020).

In line with this urgency, valid and reliable test instruments are required to accurately measure students' chemical literacy (Muntholib et al., 2020; Wiyarsi et al., 2020). Modern measurement theories, such as the Rasch model, have proven effective in objectively testing the validity of chemistry instruments (Manik et al., 2022; Mulyanti et al., 2022) as well as mapping the characteristics of contextual items (Muliani et al., 2022). Furthermore, this model is highly reliable for diagnosing critical thinking skills (Karoror & Jalmo, 2022), unraveling misconceptions (Dewi et al., 2023), and detecting students' difficulty levels (Kusnadi & Sukmawati, 2023). Rasch modeling is also robust for mapping analytical abilities and STEM literacy as a foundation for evaluating learning quality (Amsor et al., 2025; Zaky et al., 2024). Such test instruments are essential for evaluating the implementation of deep learning while serving as a baseline for more adaptive development in the future (Yerimadesi et al., 2026). Although prior research has produced chemical literacy items for topics such as acids and bases (Wahyuni & Yusmaita, 2020), atomic structure (Siti et al., 2026), thermochemistry (Afifah & Yusmaita, 2019), redox (Yuliani et al., 2021), chemical kinetics (Prasetya & Maisarah, 2024) and colligative properties of solutions (Yandriani et al., 2021), the availability of instruments specifically tailored to the buffer solution topic remains limited.

The novelty of this study lies in the development of a chemical literacy test instrument that specifically integrates buffer solution concepts with sustainable development issues, analyzed using the high objectivity of the Rasch model. Conducting this research is of paramount importance to provide teachers with a precise diagnostic measurement tool to map students' actual chemical literacy achievement. Therefore, through the application of Rasch modeling, this study aims to develop a chemical literacy test instrument for buffer solution material with comprehensive psychometric qualities to support the attainment of SDG 4 regarding quality education in senior high schools.

Method

This development research applied an instrument development model consisting of ten adaptation stages from Liu (2020). The procedural flow of all these stages is systematically presented in Figure 1. The chemical literacy test instrument was validated by five validators who are chemistry lecturers at Universitas Negeri Padang (UNP). Meanwhile, the instrument was piloted on 72 students at a public senior high school (SMAN) in West Sumatra during the even semester of the 2025/2026 academic year. The object of this study focused specifically on developing a test instrument to measure students' chemical literacy within the buffer solution topic for Phase F high school level.

Figure 1. Stages of the Rasch model analysis (Adapted and modified from Liu, 2020)

The research sequence initiated with the first stage, which involved establishing the target chemical literacy profile for Phase F students within the buffer solution material. Subsequently, the second stage was executed by formulating the construct definition, which integrated the pillars of content, context, Higher Order Learning Skills (HOLS), and attitudes. In the third stage, the instrument blueprint was constructed, and the draft items were written. The activity proceeded to the fourth stage, conducting a small-scale pilot study to gather initial student response data. These response data were then analyzed using the Rasch model in the fifth stage. In the sixth stage, the analysis focused on inspecting model fit statistics through Mean Square (MNSQ) values to examine internal consistency and item suitability. The seventh stage was carried out by utilizing variable map inspection to evaluate the alignment between the estimated item difficulty levels and the distribution of student ability levels on the same logit scale. The initial evaluation at this stage served as the baseline for conducting a large-scale field trial in the eighth stage. In the ninth stage, the final testing was performed, which included establishing construct validity claims and a comprehensive empirical examination based on unidimensionality criteria. Ultimately, the entire sequence of activities was concluded in the tenth stage by finalizing the test instrument and drafting technical guidelines for teachers. This final step serves as a tangible contribution to supporting the achievement of SDG 4 through the provision of high-quality evaluation tools in educational institutions.

Data analysis was performed using the Ministeps application. The instrument analysis components in this study included content validity testing based on expert judgment calculated through the Content Validity Index (CVI). Empirical validity (item fit) was measured based on the outfit mean square (Outfit MNSQ) criteria within the range of $0.5 \leq \text{MNSQ} \leq 1.5$, the outfit z-standard (Outfit ZSTD) within the range of $-2 \leq \text{ZSTD} \leq +2$, and the Point-Measure Correlation (Pt-Measure Corr) with an ideal criterion of $0.20 < x < 0.85$, adhering to international standards for cognitive ability tests. Instrument reliability was evaluated through the Cronbach's alpha value, alongside person reliability and item reliability indices obtained from the Rasch model's summary statistics output. The reliability coefficient criteria were classified into five categories: poor (< 0.67), fair (0.67 - 0.80), good (0.81 - 0.90), very good (0.91 - 0.94), and excellent (> 0.94).Item difficulty levels were transparently classified into four categories based on the comparison of logit measure values against the combination of the mean and standard deviation (SD) values, with the following interval definitions: very difficult if the logit value Mean ≥ +SD; difficult if Mean ≤ logit < Mean + SD ; easy if Mean ≤ -SD logit < Mean ; and very easy if the logit value < Mean – 1 SD. To objectively test and confirm whether the distribution grouping of these four difficulty categories was statistically proportional, this study utilized Linacre’s Separation Strata formula:

$$H = \frac{(G \times 4)+1}{3} \quad (1)$$

Where H represents the separation strata or the number of distinct data category groups, and G represents the separation index value obtained from the summary statistics output.

Result and Discussion

The result of this study is a chemical literacy test instrument on buffer solution material developed based on the stages of the Rasch model. To guarantee evaluation quality standards and support the achievement of SDG 4 regarding quality education, this instrument has undergone a series of empirical analyses for item quality. Based on the conducted pilot study, the evaluation results include instrument validity and reliability testing, alongside the analysis results of the difficulty index and discrimination power for each test item. The instrument analysis results from each stage are presented as follows.

Formulation of Objectives and Population

The instrument's objective is focused on measuring students' specific chemical literacy skills within the buffer solution material. The target population is specifically designated as Phase F (eleventh-grade) students of senior high schools (SMA/MA) to ensure the calibration accuracy of the Rasch model.

Construct Operationalization

This study developed a chemical literacy test instrument integrating four contextual passages with eight test items. The core components of chemical literacy embedded within these test items comprise content, context, HOLS, and affective dimensions. The instrument incorporates four varied stimulus passages linking natural ecosystem phenomena, biological homeostasis, commercial industrial applications, and clinical health aspects. The multidimensional integration across all these passages plays a pivotal role in shifting conventional assessment patterns toward authentic assessments capable of measuring students' abilities holistically and in an integrated manner (Wiyarsi et al., 2020; Sjöström et al., 2024). Collectively, the content aspect within the four passages represents the dynamic equilibrium mechanics of buffer solutions at both macroscopic and microscopic levels. Empirical findings from the evaluation of student responses demonstrate that this heterogeneous contextual approach successfully minimizes misconceptions regarding the abstract buffer material while enhancing students' memory retention (Nurhidayatulah & Prodjosantoso, 2018). Furthermore, the affective dimension within the passages proven effective in stimulating students' metacognitive awareness regarding the role of chemistry in the environment (Sjöström et al., 2024). The HOLS stimuli, provided through quantitative data presentation and substance composition tables within the passages, successfully challenged students' analytical thinking skills (Sutra & Tantra, 2020).

The analysis of student responses across the four learning objectives indicates that the majority of students have mastered the conceptual and functional levels. Students demonstrated excellent improvements in understanding theoretical and applied materials, such as foundational concepts and the functions of buffer solutions in daily life. Students' conceptual understanding emerged as the most dominant level. This result proves that contexts closely related to real life successfully assist students in understanding the mechanisms of buffer solutions logically (Nurhidayatulah & Prodjosantoso, 2018). Conversely, a shift in student ability trends was observed when dealing with pH calculation materials. Students' functional abilities heavily dominated this chemical mathematics section. This fact confirms that while students are highly proficient in applying mathematical formulas rapidly, they still encounter difficulties in explaining chemical processes at the particulate (microscopic) level when external substances are added to the solution (Shwartz et al., 2006; Zebua & Sinaga, 2021). Overall, the proportion of students at the scientific illiteracy level was remarkably low (0.8%). This finding aligns with the theory stating that scientific illiteracy occurs when students fail to respond to scientific stimuli (Shwartz et al., 2006). This success demonstrates that incorporating real-world phenomena into test passages is highly effective in helping students across all ability levels structure their reasoning when answering questions (Wiyarsi et al., 2020).

Performance Identification Results of the Defined Construct

The quantitative analysis results in Table 1 indicate that all test items meet the eligibility criteria, with CVI values ranging from 0.94 to 0.98. This index achievement is significantly above Lawshe's minimum threshold of 0.736 for a panel of 5 validators. This validity is further reinforced by the qualitative assessments from the experts; improvement suggestions from the questionnaire were comprehensively adopted to refine the linguistic aspects, stimulus clarity, and the precision of both the answer options and scoring rubrics for each item. The integration of this quantitative evaluation and qualitative refinement proves that the developed items align closely with the chemical literacy indicators and the targeted learning objectives. Concurrently, the data obtained from the experts establishes a robust theoretical foundation, demonstrating that the instrument is highly viable for use in the subsequent phases of the research (Wilson et al., 2012).

Table 1. Results of Content validity and Empirical Validity

Question items	CVI	Outfit MNSQ	Outfit ZSTD	PT measure correlation	Criteria
1.1	0.94	0.19	-0.43	0.53	Valid
1.2	0.96	0.68	-0.42	0.75	Valid
2.1	0.95	0.64	-0.37	0.77	Valid
2.2	0.95	1.36	0.72	0.57	Valid
3.1	0.96	0.68	-0.42	0.75	Valid
3.2	0.96	1.25	0.63	0.80	Valid
4.1	0.96	1.69	1.06	0.48	Valid
4.2	0.98	0.63	-0.31	0.70	Valid

Trial Results

The implementation of a small-scale trial involving 10 students generated initial empirical data in the form of student responses. Through Rasch model analysis assisted by the Ministeps application, an objective preliminary overview of the instrument's psychometric characteristics was obtained (Diyanah & Yusmaita, 2022; Yerimadesi et al., 2026). This analysis yielded several critical evaluation parameters, including item fit, difficulty level measure (logit) values, instrument reliability indices, and Wright maps (Eva & Mahmudi, 2024). Conceptually, all these empirical parameters serve to detect quality weaknesses and item variations before the instrument is deployed to a broader sample.

Conducting Rasch Model Analysis

The data in Table 1 indicate that the majority of the test items meet the model fit criteria and are declared empirically valid. The Outfit MNSQ value for item 1.1 (0.19) indicates that this item is overly predictable due to the tendentially uniform and monotonous responses from the 10 students. Conversely, the Outfit MNSQ value for item 4.1 reached 1.69. Such fluctuations in MNSQ values are frequently triggered by unexpected response patterns from students when reacting to the item stimuli (Eva & Mahmudi, 2024). All Outfit ZSTD values remained stable and hovered close to zero, even though the MNSQ values for items 1.1 and 4.1 experienced deviations. Psychometrically, the ZSTD index is a statistical transformation form that is highly sensitive to sample size (Planinic et al., 2019). Items 1.1 and 4.1 were retained and not eliminated from the instrument because both items were proven to still satisfy the other two item fit assessment criteria (Sumintono & Widhiarso, 2015).

Table 2. Reliability Results

Reliability	Small-scale	Large-scale
person reliability	0.68	0.78
item reliability	0.63	0.90
Cronbach's alpha	0.83	0.83

The Cronbach's alpha value of 0.83 and the person reliability index of 0.68 indicate that the consistency of student responses falls into the fair category. This individual stability value proves that the conceptual passage stimuli within the items were well-comprehended by the students' logical reasoning (Diyanah & Yusmaita, 2022), which constitutes a crucial aspect of designing contextual chemistry assessments. Conversely, the item reliability value of 0.63 is classified into the poor category. Respondent characteristics factored into this low index, as the trial was conducted with twelfth-grade students nearing graduation. This extended time interval triggered a decline in conceptual memory retention (decay theory) and broadened the variation of misconceptions (Haruna et al., 2024). This low item reliability reflects a high variance of measurement error inherent in small-scale testing (Siti et al., 2026). Nevertheless, for the initial stage of instrument development, this moderate figure is deemed adequate as a foundational baseline (Manik et al., 2022). Consequently, the subsequent phase proceeded with a large-scale evaluation by expanding the sample size.

Table 3. Item Measure Results

Test item	Small-scale		Large-scale	
	Item measure	Criteria	Item measure	Criteria
1.1	-2.37	Very easy	-1.60	Very easy
1.2	0.10	Difficult	0.24	Difficult
2.1	0.10	Difficult	-0.25	Easy
2.2	0.35	Difficult	0.54	Difficult
3.1	0.56	Difficult	0.69	Very difficult
3.2	1.34	Very difficult	0.51	Difficult
4.1	0.10	Difficult	-0.08	Easy
4.2	-0.19	Easy	-0.04	Easy
sD		0.99		0.68

Based on the data in Table 3, the distribution of item difficulty levels varies objectively, where item number 6 is identified as the most difficult item (1.34 logits) and item number 1 as the easiest (-2.37 logits). The higher the estimated logit value of an item, the greater the cognitive demands placed on students (Nazharita et al., 2022). This hierarchical difficulty characteristic is essential for measuring the depth of students' chemical literacy skills (Khaeruman et al., 2025). The person separation coefficient value is 1.45 with a separation strata (H = 2.26), and the item separation value is 1.32 with a separation strata (H = 2.09). Although only implemented in a small-scale trial, this instrument empirically possesses adequate sensitivity to separate student abilities into two distinct proficiency groups. The obtained item separation value proves that the construction of these eight chemical literacy items is proportionally divided into two difficulty clusters (high and low).

Review of Item Fit Statistics

All test items are empirically valid and viable to be retained within the instrument. Although indications of overfit were found in item 1.1 and unexpected response patterns were observed in item 4.1, the stable Outfit ZSTD values hovering near zero confirm that these fluctuations resulted from the constraints of the initial sample size. This characteristic is supported by the person reliability index, which falls into the fair category, proving that the conceptual passage stimuli within the items were consistently understood by the students (Diyanah & Yusmaita, 2022). Conversely, the poor item reliability coefficient is a mathematical reflection of the high measurement error typical of small-scale research (Siti et al., 2026). Nevertheless, this instrument possesses an adequate psychometric foundation, as the separation index is empirically sensitive enough to separate student abilities and classify the item difficulty structure into two proportional functional clusters. As a follow-up, minor refinements were made to the phrasing of the stimulus for items that did not yet fit to prevent comprehension bias. Subsequently, this instrument will be directly deployed in a large-scale trial with a broader sample to reduce error variance and firmly stabilize item reliability.

Review of the Wright Map

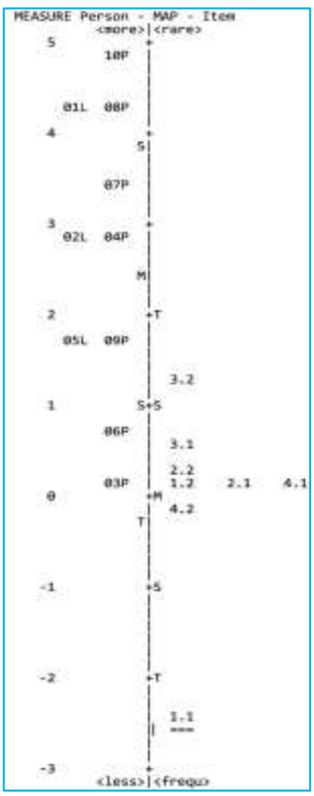

Figure 2. Wright map for the small-scale trial

The visualization of the Person-Item Map in Figure 5 indicates a tendency for the items to be overly easy, resulting from the wide gap between student abilities and item difficulty levels. The average cognitive ability of the students clusters at the upper area of the plot, whereas the average item difficulty conversely accumulates at the bottom area. This substantial empty gap proves that the current instrument is not sufficiently challenging to accurately measure the maximum ability threshold of upper-group students (Sutra & Tantra, 2020). Therefore, prior to being tested on a large scale, the instrument's quality will be enhanced by increasing the item difficulty levels to comprehensively map all levels of students' chemical literacy skills (Purwanti et al., 2022), Concurrently, this adjustment fulfills the principles of SDG 4 in providing appropriate and optimal item challenges tailored to students' intellectual potential (Ferguson & Roofe, 2020; Yerimadesi et al., 2026).

Checking Invariant Properties

A re-analysis of the instrument's invariant properties was conducted to ensure an increase in the quality of the items' empirical validity following the refinement stage. Based on the data in Table 2, item 2.1 exhibited an Outfit ZSTD value of -2.77; nevertheless, this item was retained because its Outfit MNSQ and point-biserial correlation values still satisfied the acceptable criteria (Eva & Mahmudi, 2024). This validity condition indicates a significant improvement compared to the small-scale trial, where item 1.1 was previously detected as overly predictable and item 4.1 was confusing. Theoretically, the anomalies in the initial trial were triggered by the constraints of the sample size and ambiguous phrasing (Sumintono & Widhiarso, 2015). Therefore, refinements in linguistic quality coupled with sample expansion served as an effective methodological solution to restore all test items into an accurate and ideal value range (Linacre, 2002).

Table 4. Large-Scale Empirical Validity Values

Test item	Outfit MNSQ	Outfit ZSTD	PT Measure Correlation	Criteria
1.1	1.12	0.46	0.50	Valid
1.2	1.02	0.17	0.77	Valid
2.1	0.52	-2.77	0.79	Valid
2.1	0.94	-0.28	0.72	Valid
3.1	0.75	-1.41	0.69	Valid
3.2	1.25	1.3	0.60	Valid
4.1	1.34	1.61	0.57	Valid
4.2	1.10	0.57	0.63	Valid

The results of the large-scale trial yielded a Cronbach's alpha value of 0.83 (very good), a person reliability index of 0.78 (good), and an item reliability index reaching 0.90 (excellent). This achievement marks a substantial increase from the small-scale trial, which only recorded a person reliability of 0.68 (fair) and an item reliability of 0.63 (poor). This finding demonstrates high consistency and the robust quality of the developed test items (Sumintono & Widhiarso, 2015). Theoretically, expanding the sample size and revising the item construction successfully broadened the subjects' ability range objectively, thereby enhancing and stabilizing the instrument's reliability parameters (Linacre, 2002).

Sample expansion triggered a shift in item logits toward a more objective direction. Item 2.1, which was previously categorized as difficult, shifted to the easy category, whereas item 3.1 elevated to the very difficult category. This shifting phenomenon aligns with Rasch theory, which states that an adequate sample size effectively reduces measurement bias and stabilizes the standard error values of item difficulty (Linacre, 2020). The data in Table 3 indicate that the item difficulty levels increased progressively according to conceptual complexity (Linacre, 2020). Items 1.1 and 1.2 (Basic concepts of buffer solutions) were the easiest to master because they only tested students' preliminary understanding. For items 2.1 and 2.2 (Working mechanism of buffer solutions), the difficulty levels began to vary as students were required to analyze the equilibrium shift process, which is quite abstract. The highest cognitive challenge for students was found in items 3.1 and 3.2 (pH calculation of buffer solutions), which became the most difficult item area because they integrated mathematical skills and chemical concept comprehension simultaneously. Items 4.1 and 4.2 (Applications and roles of buffer solutions) sat within a balanced difficulty range around the mean value, indicating that the contextual application concepts in daily life could be well-understood and correctly answered by students after the phrasing was refined.

Furthermore, the person separation value increased to 1.91 (separation strata H = 2.88), which is capable of dividing students into three distinct ability levels. Concurrently, the item separation value surged to 2.99 (separation strata H = 4.32), successfully separating the test items into four precise difficulty levels.

The Wright Map results in Figure 2 and Figure 3 indicate that the extreme gap between student abilities and item difficulties persists. Theoretically, this wide empty gap confirms that the test package remains generally too easy for the majority of the students; consequently, the instrument is not yet capable of accurately measuring the maximum cognitive ability threshold of the upper-group students (Sumintono & Widhiarso, 2015). The item quality refinements implemented thus far have only succeeded in decomposing the density of the item hierarchy—shifting from items that clustered uniformly at the 0-logit line in the small-scale trial to a more sequentially distributed spread in the large-scale evaluation. This shifting phenomenon aligns with the theory stating that improvements in item construction are effective in reducing comprehension bias and stabilizing the standard error values of item difficulty (Linacre, 2002). However, these refinements have not yet elevated the cognitive weight of the items into higher logit ranges (+2.00 to +5.00 logits). As a result, this instrument is empirically not yet optimal for comprehensively mapping the full spectrum of students' cognitive chemical literacy levels (Purwanti et al., 2022).

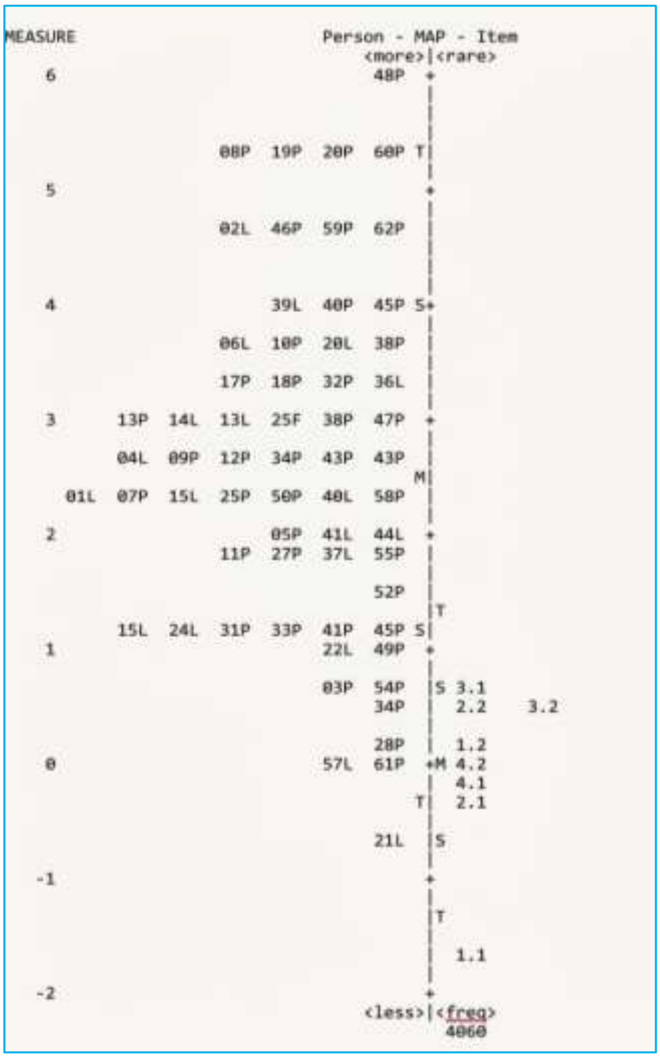

Figure 3. Wright map for the large-scale trial

Establishing Validity Claims

The integration of high content validity and misfit-free Rasch model parameters provides robust evidence that students' cognitive processes align with the scientific structure of the chemistry being tested. All items are now free from structural bias, ensuring that score variations purely represent students' chemical literacy levels. Methodologically, this claim guarantees the fairness and stability of the instrument. The instrument possesses strong psychometric validity as a standardized measurement tool, meaning the generated data holds a high degree of inference for diagnosing students' actual cognitive abilities.

Developing the Test Instrument User Guide

This study produced a final product in the form of a test instrument user guide. This manual contains detailed test administration instructions, implementation procedures, time allocation, proctoring technicalities, and a scoring system equipped with assessment rubrics for the buffer solution material. With the availability of this user guide, educational practitioners can independently administer the chemical literacy test instrument. This ease of independent replication by teachers supports the achievement of SDG 4 by expanding sustainable access to quality assessment instruments across various educational institutions.

Conclusion

Through the Rasch model, this study successfully developed a chemical literacy test instrument for high school (SMA/MA) buffer solution material that possesses valid psychometric quality along with excellent item and person separation indices. However, based on the Wright Map analysis, this instrument still exhibits limitations, as the resulting item difficulty levels tend to be relatively easy for the target population and do not yet provide a sufficient spectrum of high-logit items to assess high-achieving academic students. As an evaluation tool, this instrument functions as an objective diagnostic measure to map students' chemical literacy achievements at a specific point in time, thereby supporting the targets of SDG 4 regarding quality education. For future research development, it is recommended to modify the instrument by incorporating more complex and challenging items to reach a broader spectrum of student abilities.

Acknowledgments

The authors would like to express their gratitude to the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) for the financial support provided through the 2026 Research Funding Program. Appreciation is also extended to Universitas Negeri Padang and the Institute for Research and Community Service (LPPM) for providing the necessary facilities during this research (Contract No. 22054/UN35.15/LT/2026).

Author Contributions

S.H.Y.: Conceptualization, methodology, investigation, and original draft preparation. Y., F.G., R.: Data curation, formal analysis, and resources. Y., F.G.: Review, editing, and visualization.

Funding

This research was funded by the Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology, Fiscal Year 2026 (Contract No. 22054/UN35.15/LT/2026).

Conflicts of Interest

The authors declare no conflict of interest regarding this research.

References

Afifah, S., & Yusmaita, E. (2019). Perancangan Assesmen Literasi Kimia Pada Materi Termokimia Kelas XI SMA/MA Design of Chemical Literacy Assessment for Thermochemistry Topic for 11 th Grade Senior High School. *Edukimia Journal*, *1*(3), 79–83. Retrieved from http://edukimia.ppj.unp.ac.id/ojs/index.php/edukimia/

Amsor, A., Saputra, M. R. D., Nurdini, N., Asri, F. N., & Almaidah, S. (2025). Analytical Thinking Abilities in Physics Education Students: A Rasch-Based Diagnostic Study. *Jurnal Penelitian Pendidikan IPA*, *11*(12), 640–651. https://doi.org/10.29303/jppipa.v11i12.13444

Atta, H. B., Vlorensius, Aras, I., & Ikhsanudin. (2020). Developing an instrument for students scientific literacy. *Journal of Physics: Conference Series*, *1422*(1), 012019. https://doi.org/10.1088/1742-6596/1422/1/012019

Barus, R. A. (2024). 4C Skills of the 21St Century: Their Nature and Importance in Primary School Learning. *Multidisciplinary Indonesian Center Journal (MICJO)*, *1*(2), 689–696. https://doi.org/10.62567/micjo.v1i2.88

Dewi, Kurniawan, D. T., & Sukardi, R. R. (2023). Unravelling Misconceptions: Analising Primary School Pupils' Understanding on the Greenhouse Effect through the RASCH Model. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 9353–9362. https://doi.org/10.29303/jppipa.v9i11.5499

Diyanah, D., & Yusmaita, E. (2022). Analisis Kemampuan Literasi Kimia Siswa SMA Pembangunan Laboratorium UNP pada Topik Laju Reaksi dengan Model Rasch. *Journal Entalpi Pendidikan Kimia*, *3*(1), 26–39. https://doi.org/10.24036/epk.v3i1.245

Eva, D. kumala sari, & Mahmudi, I. (2024). *Analisis Pemodelan Rasch Pada Assessment Pendidikan (Analisis Dengan Menggunakan Aplikasi Winstep)*. PT. Pena Persada Kerta Utama.

Ferguson, T., & Roofe, C. G. (2020). SDG 4 in higher education: challenges and opportunities. *International Journal of Sustainability in Higher Education*, *21*(5), 959–975. https://doi.org/10.1108/IJSHE-12-2019-0353

Fitriyani, U., Dwi, Y., & Sunyono. (2022). Analisis Kemampuan Literasi Kimia Peserta didik SMA Negeri 6 Bandar Lampung Terhadap Tingkat Level Soal Materi Larutan Penyangga. *Jurnal Edukasi Matematika Dan Sains (JEMS)*, *10*(2), 193–204. https://doi.org/10.25273/jems.v10i2.12119

Fuadi., H., Robbia., A. Z., Jamaluddin., & Jufri, A. W. (2020). Analisis Kemampuan Literasi Sains Peserta Didik SMA Kelas X di Kota Solok. Jurnal Pendidikan Matematika dan Sains. *Jurnal Pendidikan Matematika Dan Sains*, *1*(4), 161–170. https://doi.org/10.29303/jipp.v5i2.122

Ghanem, S. (2020). E-learning in Higher Education to Achieve SDG 4: Benefits and Challenges. *2020 2nd International Sustainability and Resilience Conference: Technology and Innovation in Building Designs*. https://doi.org/10.1109/ieeeconf51154.2020.9319981

Haruna, B., Abdul Talib, C., & Quadri, A. T. (2024). Effect of Wait-Time on Attitude and Retention of Senior Secondary School Students to Learning Chemistry Concepts. *International Journal of Academic Research in Progressive Education and Development*, *13*(2), 1140–1147. https://doi.org/10.6007/ijarped/v13-i2/21327

Istyadji, M., & Sauqina. (2023). Conception of scientific literacy in the development of scientific literacy assessment tools: a systematic theoretical review. *Journal of Turkish Science Education*, *20*(2), 281–308. https://doi.org/10.36681/tused.2023.016

Karoror, I., & Jalmo, T. (2022). Profile of Critical Thinking Ability in Ecosystem Materials using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1599–1604. https://doi.org/10.29303/jppipa.v8i3.1394

Khaeruman, K., Suma, K., Maryam, S., Suastra, I., & Nurhidayati, S. (2025). Profile of Critical Thinking Skills and Cultural Literacy of Students in the Basic Chemistry Course. *Hydrogen: Jurnal Kependidikan Kimia*, *13*(1), 130–138. https://doi.org/10.33394/hjkk.v13i1.14515

Kusnadi, N. F., & Sukmawati, W. (2023). Analysis of Changes in the Level of Difficulty of Elementary School Students in Learning the RADEC Model on the Concept of Energy Transformation Using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, *9*(SpecialIssue), 1121–1127. https://doi.org/10.29303/jppipa.v9ispecialissue.4036

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch measurement transactions, 16*(2), 878.

Maisyaroh, F. Z. (2023). Pendidikan karakter dalam pembelajaran IPA abad 21. *Journal Of Natural Science Learning*, *02*(02), 75–82. Retrieved from https://jom.uin-suska.ac.id/index.php/JNSL/article/view/1307

Manik, A. C., Sutrisno, H., & Rivaldo, I. (2022). Pengujian Validitas dan Reliabilitas Instrumen Konsepsi Pembelajaran Kimia pada Masa Pandemi COVID-19 Menggunakan Pemodelan Rasch. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1561–1568. https://doi.org/10.29303/jppipa.v8i3.1486

Mayasri, A., Reza, M., & Nasir, M. (2023). Identifikasi dan Remediasi Miskonsepsi dengan Pendekatan Perubahan Konseptual pada Materi Kesetimbangan Kimia. *Jurnal Pendidikan Kimia*, *4*, 64–74. https://doi.org/10.19109/ojpk.v7i2.19810

Melati, H. A., Erlina, E., & Thongsan, N. C. (2024). Analysis of Students' Chemical Literacy Skills with Scientific Information: Case Study During Covid-19. *Jurnal Pendidikan Sains Indonesia*, *12*(1), 64–76. https://doi.org/10.24815/jpsi.v12i1.32660

Muliani, M., Novita, N., Mellyzar, M., Pasaribu, A. I., & Fadli, M. R. (2022). Analysis of the Characteristics of the Ethnoscience-Based Numeracy Test Instrument Using the Rasch Model. *Jurnal Penelitian Pendidikan IPA*, *8*(5), 2176–2183. https://doi.org/10.29303/jppipa.v8i5.2285

Mulyanti, S., Suwahono, S., Setiowati, H., & Ningrum, L. S. (2022). Validity Analysis Using the Rasch Model in the Development of Alkane Concept Test Instruments. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1142–1147. https://doi.org/10.29303/jppipa.v8i3.1383

Muntholib, Ibnu, S., Rahayu, S., Fajaroh, F., Kusairi, S., & Kuswandi, B. (2020). Chemical literacy: Performance of first year chemistry students on chemical kinetics. *Indonesian Journal of Chemistry*, *20*(2), 468–482. https://doi.org/10.22146/ijc.43651

Mustika, A., Hernai, & Mudzakir, A. (2020). Analisis Muatan Literasi Sains Pada Buku Teks Kimia SMA di Kota Bandung. *Jurnal Riset Dan Praktik Pendidikan Kimia*, *11*(2), 39–49. https://doi.org/10.17509/jrppk.v8i2.52291

Mutiah, M., & Siahaan, J. (2022). Identifikasi Tingkat Literasi Kimia-Sma Mahasiswa Progdi. Pendidikan Kimia Fkip Unram. *Chemistry Education Practice*, *5*(2), 151–156. https://doi.org/10.29303/cep.v5i2.3935

Nazharita, E., Hidayat, R., & Rahmatudin. (2022). Domain Geometri Siswa Smp Menggunakan Rasch. *Pedagogy*, *10*(3), 989–1000. https://doi.org/10.30605/pedagogy.v10i3.6785

Nurhidayatulah, N., & Prodjosantoso, A. K. (2018). Miskonsepsi materi larutan penyangga. *Jurnal Inovasi Pendidikan IPA*, *4*(1), 41–51. https://doi.org/10.21831/jipi.v4i1.10029

OECD. (2023). PISA 2022 Results The State of Learning and Equity in Education. In *Factsheets: I*. Retrieved from https://www.oecd-ilibrary.org/educstion/pisa-2022-results-volume-i_53f23881-en

Pertiwi, U. D., Atanti, R. D., & Ismawati, R. (2018). Pertiwi 2018, pentingnya literasi sains pada pembelajaran IPA. *Indonesian Journal of Natural Science Education*, *1*(1), 25–29. https://doi.org/10.31002/nse.v1i1.173

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Prasetya, C., & Maisarah. (2024). The Effect of Chemistry Teaching Materials on Scientific Literacy and Critical Reasoning in SMA Negeri Langsa. *IJCER (International Journal of Chemistry Education Research)*, *8*(1), 1–6. https://doi.org/10.20885/ijcer.vol8.iss1.art1

Purwanti, A. D., Mulyani, B., & Ulfa, M. (2022). Analisis Muatan Literasi Sains Dan Higher Order Thinking Skills Dalam Buku Ajar Kimia Sma Kelas X Pada Materi Reaksi Redoks. *Jurnal Pendidikan Kimia*, *11*(2), 177–123. https://doi.org/10.20961/jpkim.v11i2.54817

Putra, W. T. G. (2022). Asesmen Literasi Sains Tema Kimia Hijau Pada Siswa. *Indonesian Journal of Educational Development*, *3*(1), 1–15. https://doi.org/10.5281/zenodo.6557263

Rahmasari, U. D., Nugraha, A., & Abdul, M. (2022). Persepsi Guru Mengenai Pentingnya Kemampuan Mengembangkan Soal Tes Berbasis Literasi Dan Numerasi Di Sekolah Dasar. *COLLASE (Creative of Learning Students Elementary Education)*, *5*(6), 1105–1112. https://doi.org/10.22460/collase.v5i6.12345

Sariati, N. K., Suardana, N., & Wiratini, N. M. (2020). Analisis Kesulitan Belajar Kimia Siswa Kelas XI pada Materi Larutan Penyangga. *Jurnal Imiah Pendidikan Dan Pembelajaran*, *4*(April 2020), 76–87. https://doi.org/10.23887/jipp.v4i1.15469

Shwartz, Y., Ben-Zvi, R., & Hofstein, A. (2006). The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students. *Chemistry Education Research and Practice*, *7*(4), 203–225. https://doi.org/10.1039/B6RP90011A

Siska, H. Y., & Yerimadesi, Y. (2023). Validitas dan Praktikalitas Modul Reaksi Kimia Berbasis Guided Discovery Learning Terintegrasi Etnosains untuk Fase E SMA. *Jurnal Pendidikan MIPA*, *13*(2), 1–7. https://doi.org/10.37630/jpm.v13i2.986

Siti, F., Yerimadesi, Yuhelman, N., & Iryani. (2026). Design of a Chemical Literacy Test Instrument on Atomic Structure Material for Senior High School Students. *Jurnal Penelitian Pendidikan IPA.*, *21*(1), 164–170. https://doi.org/10.29303/jpm.v21i1.11447

Sjöström, J., Yavuzkaya, M., Guerrero, G., & Eilks, I. (2024). Critical Chemical Literacy as a Main Goal of Chemistry Education Aiming for Climate Empowerment and Agency. *Journal of Chemical Education*, *101*(10), 4189–4195. https://doi.org/10.1021/acs.jchemed.4c00452

Sumarni, W., Sumarti, S. S., & Kadarwati, S. (2023). Blended Inquiry Learning With Ethno-Stem Approach for First-Semester Students' Chemical Literacy. *Jurnal Pendidikan IPA Indonesia*, *12*(3), 439–450. https://doi.org/10.15294/jpii.v12i3.45879

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment pendidikan*. Trim komunikata.

Sutra, R., & Tantra, D. (2020). Students' Reading Competency on High Order Thinking Items in The Junior High Schools. *Journal of Education Research and Evaluation*, *4*(4), 380. https://doi.org/10.23887/jere.v4i4.28011

Wahyuni, A., & Yusmaita, E. (2020). Perancangan Instrumen Tes Literasi Kimia Pada Materi Asam dan Basa Kelas XI SMA/MA. *Edukimia*, *2*(3), 106–111. https://doi.org/10.24036/ekj.v2.i3.a186

Widya Kusuma, K., Larasyati, W., & Ventista Risma, F. (2022). Pentingnya Penerapan Budaya Literasi Membaca dan Menulis Terhadap Prestasi Peserta Didik Di Sekolah Dasar Meyla. *Jurnal Pendidikan Dan Konseling*, *4*(20), 2036–2039. https://doi.org/10.31004/jpdk.v4i4.5180

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/0748175612440286

Wiyarsi, A., Pratomo, H., & Priyambodo, E. (2020). Vocational high school students' chemical literacy on context-based learning: A case of petroleum topic. *Journal of Turkish Science Education*, *17*(1), 147–161. https://doi.org/10.36681/tused.2020.18

Yandriani, Rery, R. U., & Erna, M. (2021). Developing and Validating the Assessment Instruments to Measure Students' Analytical Thinking Ability and Chemical Literacy on Colligative Properties. *Journal of Physics: Conference Series*, *1788*(1).

https://doi.org/10.1088/1742-6596/1788/1/012027
Yerimadesi, Iryani, Faizah, Q. A., Gazali1, F., Nofri, Y., Rivo, J. S., Salsabila1, D., Alifa1, Z. Z., & Mentari, A. Z. (2026). Development of a Deep Learning Assessment Instrument in Senior High School Chemistry Education Using the Rasch Model to Support SDG 4 : Quality Education. *Jurnal Penelitian Pendidikan IPA.*, *12*(4), 626–638. https://doi.org/10.29303/jppipa.v12i4.14841
Yerimadesi, Y., Bayharti, B., Azizah, A., Lufri, L., Andromeda, A., & Guspatni, G. (2019). Effectiveness of acid-base modules based on guided discovery learning for increasing critical thinking skills and learning outcomes of senior high school student. *Journal of Physics: Conference Series*, *1185*(1). https://doi.org/10.1088/1742-6596/1185/1/012151
Yerimadesi, Y., Desi Lisa, R., Andromeda, A., & Guspatni, G. (2023). The effectiveness of guided discovery learning-based salt hydrolysis e-module in improving higher-order thinking skills (HOTS). *AIP Coference Proceeding*, *2673*(1). https://doi.org/10.1063/5.0127089
Yuliani, F., Yusmaita, E., & Gazali, F. (2021). Pengembangan Instrumen Literasi Kimia pada Tema Silver Tarnish dan Polishing Silver. *Edukimia*, *3*(1), 091–097. https://doi.org/10.24036/ekj.v3.i1.a247
Yusmar, F., & Fadilah, R. E. (2023). Analisis Rendahnya Literasi Sains Peserta Didik Indonesia: Hasil Pisa Dan Faktor Penyebab. *LENSA (Lentera Sains): Jurnal Pendidikan IPA*, *13*(1), 11–19. https://doi.org/10.24929/lensa.v13i1.283
Zaky, M., Jamhari, M., Supriyatma, Muslimin, Jamhari1, M., & Ratman. (2024). STEM Literacy Proficiency in Prospective Physics Educator: A Comprehensive Analysis Using Rasch Measurement Theory. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 10799–10810. https://doi.org/10.29303/jppipa.v10i12.8970
Zebua, E. A., & Sinaga, K. (2021). Developing Senior High School Student's Chemical Literacy Skills Through Inquiry-Based Online Chemistry Learning. *Jurnal Akademika Kimia*, *10*(2), 126–132. https://doi.org/10.22487/j24775185.2021.v10.i2.pp126-132