Assessing Chemical Literacy of Pre-service Chemistry Teacher: Rasch Analysis

Marfuatun Marfuatun[1,3], *Nahadi* Nahadi[2,*], *Galuh* Yuliani[2], *Hernani* Hernani[2]

[1]Department of Natural Sciences Education, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, Indonesia;
[2]Department of Chemistry Education, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3]Department of Chemistry Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Abstract. The study aimed to validate the chemical literacy (CL) test instrument for pre-service chemistry teachers (PCTs), categorize the CL-item difficulty, and evaluate PCTs' chemical literacy levels. The participants were 49 students in the chemistry education department at Daerah Istimewa Yogyakarta. Three different scientific knowledge were employed to assess student chemical literacy. A paper-based test was carried out for the data collection. The findings revealed that the developed CL test achieved both validity and reliability in its assessments. Interestingly, while a majority of the test items were categorized as relatively easy, the students exhibited a lower-than-expected proficiency in completing the CL test.

1 Introduction

Chemical literacy (CL) is the ability of students to comprehend and analyze daily phenomena using their knowledge of chemistry. Literacy is needed to foster meaningful study in a discipline [1]. In chemistry, literacy in chemistry helps students to understand, synthesize, and communicate chemical content knowledge [2]. They are trained to understand how chemistry is applied and how chemistry affects human life. CL is also intended to help people to anticipate future problems that can not be predicted. The future challenges include issues related to scientific misinformation [3], and sustainability development anchored on Environmental, Economic, and Equity pillars [4]. Through CL, students are empowered to make informed decisions and take responsible actions based on scientific principles.

Furthermore, students' mastery of chemical literacy needs to be evaluated continuously to determine the extent to which the student has fulfilled the skills of the domain of CL. The domains reflect different philosophies about the goals and outcomes of chemical education in diverse societies [5]. In the testing process, the domains are articulated in a framework that draws the construct to be measured, and the links between

* Corresponding author: nahadi@upi.edu

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

the construct and the content of the instrument [6]. The framework refers to guidelines for developing test instruments to measure students' CL. The most widely used CL domain is the framework proposed by Shwartz et al.[7], which includes four dimensions; chemistry content knowledge, chemistry context, high-order learning skills, and affective [8]. Another framework often used is a derivative of the PISA scientific literacy which also has four domains; knowledge, context, competencies, and attitude [9].

Many studies have developed test instruments to measure CL, but few have been proposed for pre-service chemistry teachers (PCTs). In the timespan 2015-2023, only four papers reported the process of developing CL test instruments PCTs [10], [11], [12], [13]. While, in the future, they will take on the role of actors in the learning transformation that provides basic knowledge and problem-solving skills to improve CL level. Moreover, the previous study finds a lack of teacher readiness in supporting students' scientific literacy development [14], which correlates with CL exploration. Based on a preliminary study of the exam documents for pre-service chemistry teachers from four Public Universities, the test instruments used by lecturers have not accommodated all aspects of chemical literacy. Most of the tests emphasize the chemistry content domain. Therefore, there is still a need to develop a comprehensive CL-test instrument.

This study developed the CL-test instrument based on the science literacy framework proposed by PISA 2025. Science, sustainability, and technology discourses are central themes in the framework. PISA 2025 refines and extends the PISA 2015 competencies. The competencies include explaining scientific phenomena, constructing and evaluating designs for scientific inquiry, and critically interpreting scientific data and evidence. They also involve researching, evaluating, and utilizing scientific information for informed decision-making and action [15]. In PISA 2025, students will dive into the world of misinformation, often referred to as "fake news," by exploring how scientists build a foundation of evidence-based knowledge. The instrument developed refers to the content of the General Chemistry course. The course covers many topics [16] and is more easily developed for studying phenomena through an interdisciplinary approach [17].

One of the crucial parts of developing test instruments is related to the quality of the instrument, which is validity and reliability. Both parameters can be analyzed using the Rasch model. Rasch analysis helps researchers construct instruments, monitor instrument quality, and calculate respondent performance accurately [18], even in small samples [19]. Rasch analysis can maximize trait homogeneity and determine the true abilities of students [20]. Therefore, Rasch analysis was utilized in this study to measure the validity and reliability of the CL-instrument test and evaluate CL levels of PCTs. The following three research questions were formulated : (1) What are the reliability and validity of the CL instrument test based on the PISA 2025 framework for PCTs, as assessed using Rasch parameters? (2) What is the categorization of the difficulty of CL items; and (3) What is the classification of the PCTs' CL level?

2 Method

This study utilized a cross-sectional research design with a quantitative approach, conducted in June 2024. The participants are in the 2nd year of PCTs in Daerah Istimewa Yogyakarta collected by convenience sampling and consisted of 49 students, of whom 6 were males (12.24%) and 43 were females (87.76%). A detailed characteristic of the participants is shown in Table 1. Participants received CL questions in a paper-based exam, with 100 minutes allotted to complete the proctored test.

Table 1. Demographic of Participant

Participants	**Frequency**	**Percentage (%)**
Gender		
Female	43	87.76
Male	6	12.24
General Chemistry Grade		
A	11	22.45
B	35	71.43
C	3	6.12
GPA honors		
Cumlaude (C), GPA ≥ 3.51	23	46.94
Second Class (S), GPA<3.51	26	53.06

The PISA 2025 science framework was adapted to develop CL questions in this study. The instrument test consists of 50 multiple-choice questions with five possible answers. The questions were categorized into three sub-tests which are types of scientific knowledge; content knowledge (C), procedural knowledge (P), and epistemic knowledge (E). The topics of the CL questions were developed based on general chemistry handbooks used in pre-service chemistry teacher education at six public universities in Indonesia. The discourses presented in the questions consist of ten issues related to misinformation and sustainable development. The test instrument had been content validated by experts with Ave-CVI = 0.97

The data was analyzed by the Rasch model, carried out by WINSTEP 5.2.4 software. The Rasch model is capable of determining the item difficulty independent of test takers, and the person's ability independent of test items, thereby providing additional precision and confidence in the student scores on the multiple-choice test [21]. The assessment of the CL-instrument test's validity and reliability was conducted through the application of Rasch measurement parameters. This evaluation involved an examination of unidimensionality, which ensures that the test measures a single construct effectively. Additionally, we rigorously evaluated both person and item reliability criteria. The Wright map was presented to perform students' actual abilities. However, DIF is not calculated because sample sizes are less than 200 [22]. Furthermore, logit value item (LVI) and logit value person (LVP) analysis were performed to classify item difficulty and student ability.

3 Result and Discussion

Based on the research objective, the findings are discussed under the following themes: validity of CL-test instrument reliability of instruments test, DIF across students' general chemistry grades and cumulative grade points, the evaluation of pre-service chemistry teachers' CL level, and item difficulties categories.

3.1 Validity and Reliability of Chemical Literacy Test

3.1.1. *Validity*

In this study, we evaluated whether the sub-tests (C, P, and E) followed the basic assumption of measurement, unidimensional. Two criteria were used for judging unidimensionality; (1) the variance explained by the measurement dimension be at least 20% [23], and (2) the variance unexplained of the residuals is no more than 15% [24]. Both values for all sub-tests are shown in Table 2. The result revealed that all sub-tests of CL had raw variance ranging from 23% to 40% which indicates the CL test achieved an acceptable threshold for unidimensional. The unidimensionality of the test is reinforced by the unexplained variance whose maximum value is 10.2%. The values indicate that there were no substantial and meaningful secondary dimensions in the data.

Two parameters were used to evaluate the validity of the CL-test instrument; outfit mean squares (MNSQ) and z-standardized (ZSTD), both of which are identified for items and persons. The range of outfit MNSQ used in this study was 0.5 – 1.5, the value range indicates that the developed instrument is productive for measurement [25]. Acceptable values for ZSTD are between -2.0 and +2.0 with sample sizes between 30 and 300 [26]. The ZSTD value can be ignored if the MNSQ value is within the acceptable fit range [25]. The results of the Rasch analysis (Table 2) revealed that the developed Cl test meets the validity criteria according to the Rasch parameters for the entire test and each scientific knowledge. In addition, each knowledge in the CL test is the ideal value for fit criteria of test items, with 1.00 logits for MNSQ and 0.0 for ZSTD [27].

Table 2. Summary of the Rasch Analysis for the chemical literacy test and each sub-test

Psychometrics attribute	Scientific Knowledge			CL test
	C	P	E	
Number of items	23	11	16	50
Mean				
Item outfit MNSQ	1.00	1.23	1.08	1.07
Person outfit MNSQ	1.00	1.23	1.08	1.07
Item outfit ZSTD	-0.1	0.3	-0.1	-0.1
Person outfit ZSTD	0.1	0.3	0.1	0.1
Unidimensionality				
Raw variance explained by measures	23.2%	40.8%	27.0%	
Unexplained variance in 1st contrast	10.2%	9.1%	9.4%	

3.1.1. *Reliability*

The reliability was judged following two indicators, person and item reliability [28], [29], and Croncbach's alpha (α) [30]. The reliability results for all sub-tests are summarized in Table 3. The item reliability values ranged from 0.71 to 0.90, while person reliability just achieved ranging from 0.60 to 0.86 for all sub-tests as well as the entire test. The result pointed out that the quality of questions in the CL test had good reliability, but the consistency of students' answers was poor. In addition, Croncbach's alpha value ranged from 0.65-0.88 indicating adequate reliability [30]. Overall, the CL test and all its sub-tests met the reliability criteria for the Rasch parameters.

Table 3. Reliability results

Reliability	Instruments			
	C	P	E	CL Test
Item reliability	0.71	0.90	0.85	0.84
Person reliability	0.80	0.71	0.60	0.86
Croncbach's alpha (α)	0.74	0.78	0.65	0.88

3.2 Item Difficulty Categorization of The Chemical Literacy Test

In Rasch analysis, item difficulty can be examined by referring to item separation (G) and a "logit" scale that extends from negative infinity to positive infinity [18]. The finding in Table 4 presents the CL test and the sub-tests met the item separation threshold which is 2 logits [31], except procedural knowledge which just reached 1.55 logits. However, the number of item separation strata ($H = (4G + 1)/3$) remains more than 2, so the item difficulty levels can be classified into three categories; easy, moderate, and hard.

Table 4. Summary of Item and Person Measure

Attribute	Sub-tests			CL test
	C	P	E	
Item separation	2.00	1.55	2.37	2.28
Item Measure (logit)				
Mean	0.00	0.00	0.00	0.00
SD, standard deviation	0.79	1.27	0.92	0.88
SE, standard error	0.17	0.40	0.24	0.13
Person separation	1.56	1.43	1.23	2.49
Person Measure (logit)				
Mean	-0.55	-0.19	-0.35	-0.39
SD, standard deviation	0.96	1.76	0.97	0.91
SE, standard error	0.14	0.25	0.14	0.13

Furthermore, the items were classified by dividing the distribution of the mean logits of items (0.00 logits) and SD (0.88) [32]. Using these thresholds, the CL test was classified under the categorization; hard (LVI $\geq$ 0.88), moderate (0.88 > LVI $\geq$ 0.00), and easy (LVI < 0.00). The classification of CL item difficulty is presented in Table 5. The results of the LVI analysis revealed that the majority of item tests are easy for PCTs in each scientific knowledge group ranging from 47% to 63%. It is noteworthy that the proportion of difficult items for all subtests was less than 20%. One can assume that students have a good level of chemical literacy in all aspects of scientific knowledge.

Table 5. Categorization of Chemical Literacy Item Difficulty

Sub-test	**Difficulty level**		
	Hard (LVI ≥ 0.88)	**Moderate (0.88 > LVI ≥ 0.00),**	**Easy (LVI < 0.00)**
C	C1, C3	C10, C12, C19, C2, C6, C13, C17, C20, C21, C14	C18, C5, C15, C11, C16, C22 C4, C7, C8, C9, C23
P	P6	P3, P7, P2	P8, P11, P1, P4, P9, P5, P10
E	E11, E5, E1	E6, E8, E10	E2, E7, E3, E14, E15, E16, E4, E12, E9, E13

The difficulty level can be confirmed by the Wright map representing the interaction between items and persons (Fig. 1). The Wright map reveals that more than half of the items are below the mean item logit scale. Item P6 (2.37 logits) was indicated as the most difficult item, while C18 (1.84 logits) was the easiest item. The results further demonstrate that the study follows previous analyses, the item difficulty level ranges from −3 logits to +3 logits [18].

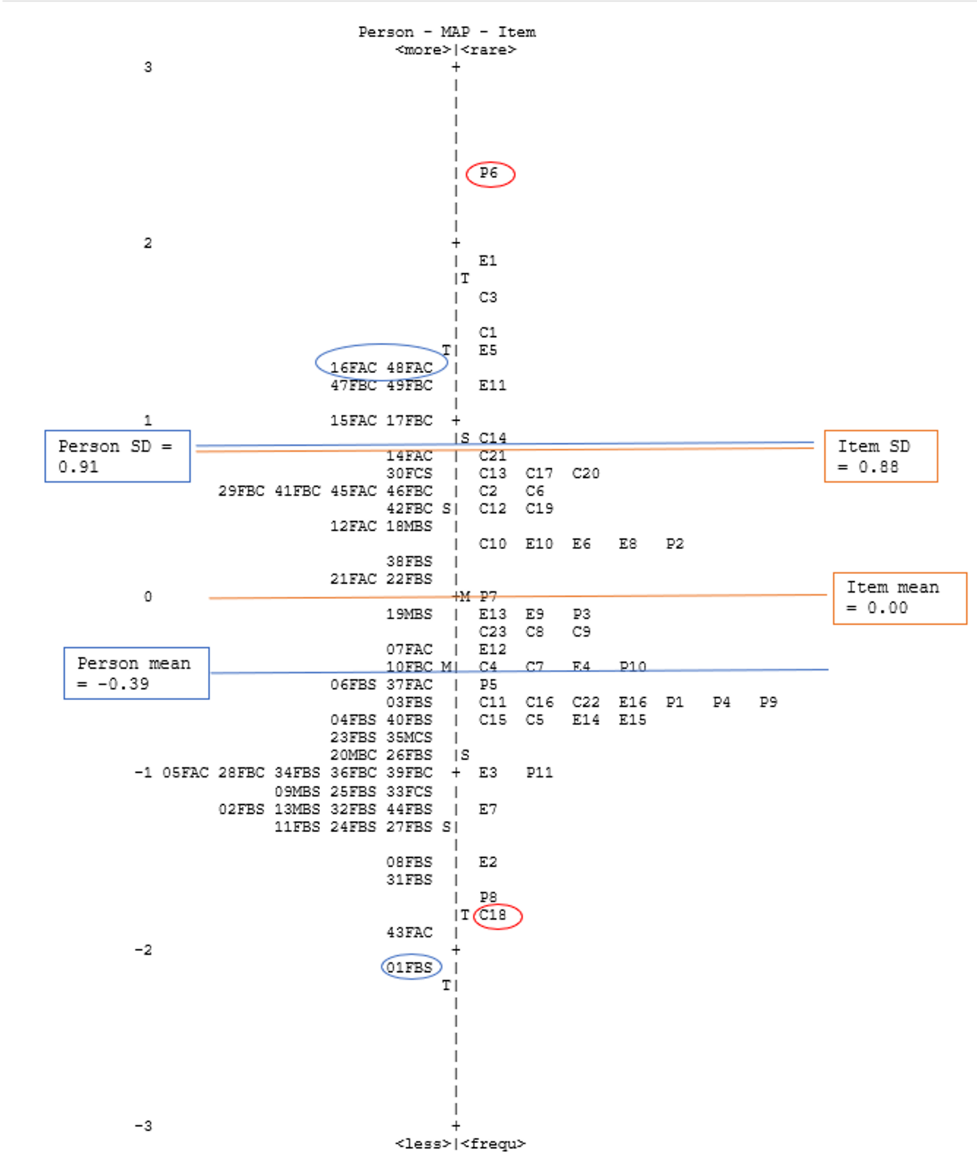

Fig. 1. The Wright map representing the interaction between items and persons

3.3 The evaluation of Pre-service Chemistry Teachers' Chemical Literacy

The students' chemical literacy was evaluated based on their ability to solve problems on the overall scientific knowledge. Student ability was determined by comparing the mean person measure of the sub-test with the mean item measure whereby 0.00 logits is a default value [18]. According to Table 3, student performance in solving the items of the CL test was below average, as indicated by negative values in all sub-tests (C -0.55, P = -0.19, E= -0.35). This finding was supported by the Wright map (Fig. 1) that more than half of the students are below the mean item logits. In general, PCTs' ability in the test is lower than the difficulty of the item. Additionally, the contribution sub-test scores on the CL scores were assessed by calculating the correlation coefficient for all sub-tests and the overall test

(Table 6). The results showed that all correlation coefficients were positive and significant, ranging from 0.60 to 0.91. The highest correlation was observed between the content knowledge and CL test (r = 0.91), and the lowest correlation was found between the procedural and content as well as epistemic. The finding hinted that PCTs with higher scores on the sub-test would achieve a higher score on the CL test.

Table 6. Correlation between the sub-test and the CL test

Test-Subscale	Logit Min	Logit Max	Pearson Correlation			
			C	P	E	CL-test
C	-3.37	1.17	1			
P	-4.07	4.65	0.60	1		
E	-2.22	1.71	0.70	0.62	1	
CL test	-2.06	1.32	0.91	0.81	0.88	1

Furthermore, the ability of PCTs to solve the CL test was classified through LVP analysis according to the mean logits of the person (-0.39) and SD (0.91), the complete person measure is shown in Table 3. Using these thresholds, the PCTs' ability was classified under the categorization; high (LVP ≥ 0.91), moderate (0.91 > LVP ≥ -0.39), and easy (LVP < -0.39). The results revealed that 6 (12%) students were categorized as having high abilities, 15 students (31%) were categorized as possessing moderate abilities, and 28 students (57%) had low abilities. There are two students with the highest abilities, both of whom are female, having an "A" grade in the general chemistry course, and Cumlaude honors. While the lowest ability was a female student, having a "B" grade in the general chemistry course, and second-class honors.

In general, most PCTs had low abilities in solving the CL test and the findings seem to support the previous study [10], [12] that students are still having difficulties in solving the CL test. The study informs that student achievement in the general chemistry course will affect their ability to solve questions in the CL test, just like their success in other courses as indicated by the GPA honors. Therefore, general chemistry grade and GPA could be used as predictors for PCTs in CL measurement.

4 Conclusions

The results of the study show that the developed CL test has been proven to be valid and reliable, thus it can be used to measure the CL level of students in chemistry teacher education. The grouping of the difficulty items informed a wide range of difficulty levels, where the distribution of difficulty levels is balanced between the three scientific knowledge; content, procedural, and epistemic. Most of the PCTs were classified as low abilities, in which their CL scores are the result of equal contributions between scientific knowledge. Therefore, we suggest that this chemistry literacy should be more intensively embedded and trained in chemistry teacher education, thus the students become individuals who are literate in the chemistry discipline. In addition, the small sample size is a limitation of our study, so more participants are needed to validate our findings.

References

1. Z. Kohen, O. Herscovitz, Y. J. Dori, Chem. Educ. Res. Pract., **21**, 1 (2020)
2. State Government of Victoria, Accessed: Oct. 26 (2021), [Online], Available: https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/english/literacy/Pages/introduction_to_literacy_in_science.aspx
3. J. Osborne, D. Allchin, Int J Sci Educ (2024)

4. S. Mitchell, A. J. Martín, G. Guillén-Gosálbez, J. Pérez-Ramírez, Angew. Chem. Int. Ed., **63**, 6 (2024)
5. P. W. U. Chinn, Cult Stud Sci Educ, **7**, 2 (2012)
6. J. Pearce, D. Edwards, J. Fraillon, H. Coates, B. J. Canny, D. Wilkinson, Perspect Med Educ., **4**, 3 (2015)
7. M. Marfuatun, N. Nahadi, G. Yuliani, H. Hernani, JPPIPA,**10**, 6 (2024)
8. Y. Shwartz, R. Ben-Zvi, A. Hofstein, Chem. Educ. Res. Pract., **7**, 4 (2006)
9. OECD, *PISA 2015 Results (Volume I): Excellence and Equity in Education* (OECD, Paris, 2015)
10. W. Sumarni, Sudarmin, Wiyanto, A. Rusilowati, E. Susilaningsih, TUSED, **14**, 3 (2017)
11. E. I. Nada, W. K. Sari, IOP Conf Ser Earth Environ Sci, **1796**, 1 (2021)
12. M. Muntholib, S. Ibnu, S. Rahayu, F. Fajaroh, S. Kusairi, B. Kuswandi, Indonesian Journal of Chemistry, **20**, 2 (2020)
13. D. Fadly, S. Rahayu, I. W. Dasna, Y. Yahmin, International Journal of Instruction, **15**, 1 (2022)
14. J. Siswanto, S. Suyidno, S. Mahtari, F. Fitriyani, W. Febriani, E. Sari, JPSI, **11**, 2 (2023)
15. OECD, *PISA 2025 Science Framework (Draft)* Accessed: Dec. 21 (2023) [Online]. Available: https://pisa-framework.oecd.org/science-2025/
16. M. M. Cooper, R. L. Stowe, Chem Rev, **118**, 12 (2018)
17. M. Balabanoff, H. al Fulaiti, B. DeKorver, M. Mack, A. Moon, Chem. Educ. Res. Pract., **23**, 2 (2021)
18. W. J. Boone, CBE Life Sci Educ, **15**, 4 (2016)
19. B. Babcock, K. J. Hodge, Educ Psychol Meas, **80**, 3 (2020)
20. A. Darmana, A. Sutiani, H. A. Nasution, I. Ismanisa, N. Nurhaswinda, JPSI, **9**, 3(2021)
21. W. J. Boone, K. Scantlebury, Sci Educ, **90**, 2 (2006)
22. J. Blackwood, S. Carpentier, W. Deng, A. Van De Winckel, PLoS One, **18**, 6 (2023)
23. M. D. Reckase, Journal of Educational Statistics, 4, 31(979)
24. B. Sumintono, W. Widhiarso, *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. (Trim Komunikata, Cimahi, 2014)
25. W. J. Boonee·, J. R. Staver, M. S. Yale, *Rasch Analysis in the Human Sciences* (Springer, Dordrecht, 2014)
26. T. G. Bond, Z. Yan, M. Heene, *Applying the Rasch model: Fundamental measurement in the human sciences* (Routledge, New York, 2020)
27. B. Sumintono, W. Widhiarso, *Aplikasi pemodelan Rasch pada assessment pendidikan*. (Trim komunikata, Cimahi, 2015)
28. W. P. Fisher, Rasch Measurement Transactions, **21**,1 (2007)
29. S. Soeharto, B. Csapó, Thinking Skills and Creativity, **46** (2022)
30. K. S. Taber, Res Sci Educ, 48, 6 (2018)
31. T. G. Bond, C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Routledge, New York, 2015)
32. D. Adams, M. Tan Hwee Joo, B. Sumintono, O. Siew Pei, MJLI, **17,** 1 (2020)