

Analysis of eleventh-grade students' scientific literacy based on PISA-framework in reaction rate topic

Muntholib Muntholib^{1,1}, Levina Ingrid Al Maudrey¹, Munzil Munzil¹, M. Muchson¹, Nur Candra Eka Setiawan¹

¹Chemistry Department, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, 65145, East Java, Indonesia.

Abstract. Mapping domain-level scientific literacy can guide instruction on specific chemistry topics. Objective: To profile eleventh-graders' scientific literacy on reaction rate using the PISA 2015 science framework (context, scientific competencies, scientific knowledge). Methods: Cross-sectional descriptive survey with N = 102 students (three intact classes). A 35-item multiple-choice test was developed; 25 items that met the point-biserial validity criterion (≥ 0.30) and yielded $\alpha = 0.87$ were retained in the final instrument. Scores (total and domains) were expressed as percent-correct; analyses were descriptive only (mean, SD, median, IQR, range). Results: Overall score reached M = 75.29%, SD = 20.65, median 84%, IQR 57–96%, range 24–100%. Domain means were 75.17% for context, 76.12% for scientific competencies, and 73.47% for scientific knowledge. Conclusion: Findings indicate mid-to-high performance with a relatively balanced domain profile and scope to strengthen scientific knowledge. These percentages are not directly comparable to PISA ability-scale scores (different metric and coverage); results are intended for internal mapping to inform improvement in reaction-rate instruction.

1 Introduction

Scientific literacy is now a main goal of modern science education because it helps people understand scientific information, think logically about evidence, and make good choices in their own lives and in society. It encompasses not only the mastery of scientific facts but also the ability to apply concepts, comprehend scientific inquiry, and assess data-driven assertions [1–3]. The PISA 2015 science framework defines scientific literacy through three interconnected domains: context, scientific competencies, and scientific knowledge [1]. These areas include students' ability to use scientific concepts in real-life situations, to think about problems in ways that lead to inquiry, and to use content, procedural, and epistemic knowledge to support scientific explanations.

Indonesia's scores in PISA 2018 and PISA 2022 indicate that scientific literacy remains a challenge. The national average scores are still below the OECD average, and only a small number of students are proficient at Level 2 or higher [2,3]. These results show that there are problems with the system, but it's important to remember that PISA reports ability-scale scores

¹ Corresponding author: muntholib.fmipa@um.ac.id

based on item-response theory, while classroom tests usually report percent-correct scores on limited sets of topics. You can't compare these two numbers directly, and you shouldn't treat school-level percentages as equivalent to PISA scale scores [1–3].

The topic of reaction rates in chemistry classes has significant potential to improve students' scientific literacy, as it prompts them to think about how things change over time and to connect chemical principles to real-world phenomena [4]. But students' understanding of concepts and their ability to reason from evidence are still hampered by persistent misunderstandings of collision theory, temperature effects, concentration, surface area, and the role of catalysts [4–6]. Previous research has indicated that students in secondary and tertiary education often encounter challenges in chemical kinetics reasoning and contextual problem-solving [14]. These misunderstandings show how important it is to have tests that measure not only what you know but also how well you can analyze data and judge explanations. Inquiry-based learning (IBL) is a well-established and effective way to teach chemistry, helping students become more scientifically literate. Systematic reviews and meta-analyses show that guided inquiry is especially good at helping students improve their reasoning, understanding of concepts, and ability to use scientific evidence [7,14]. When used to teach reaction rates, IBL can prompt students to ask questions, plan investigations, analyze graphs, and make sense of patterns in real life. Previous research in chemistry education shows that well-designed inquiry tasks can help students understand chemistry better and think more critically [4,7,12].

Even with these improvements, few studies have used the full structure of the PISA 2015 framework to map scientific literacy at the domain level for a specific chemistry topic. Most research focuses only on general scientific literacy or conceptual mastery. It doesn't look at how well students do in all three areas: context, competency, and knowledge, in a way that is in line with the PISA standards. Furthermore, insufficient attention has been paid to evaluations in high-performing public schools, where inquiry-based pedagogical approaches may foster competencies in some areas while leaving enduring deficiencies in others.

Consequently, this study aims to provide a descriptive, PISA-aligned domain profile of Grade 11 students' scientific literacy regarding the reaction-rate topic at a high-achieving Indonesian senior high school. This study examines how well students perform in the areas of context, scientific knowledge, and scientific competencies using a validated multiple-choice tool based on the PISA 2015 science framework. It also gives detailed descriptive statistics for the whole class and each individual student. The study adds to a more nuanced understanding of domain-specific scientific literacy in chemistry, and evidence to help with inquiry-based instructional design. It also makes it clear that percent-correct scores from this study should not be directly compared to PISA ability-scale scores.

2 Method

2.1 Research design

This study employed a descriptive survey design to profile students' scientific literacy regarding the reaction-rate topic, according to the PISA 2015 science framework. Because the purpose of this research was to obtain a domain-level description rather than to test causal hypotheses, no experimental manipulation, group comparison, or inferential statistical testing was conducted. All analyses were descriptive and focused on mapping student performance across the domains of context, scientific competencies, and scientific knowledge.

2.2 Participants and setting

The three classes comprised 34 students each ($N = 102$ in total) of Grade 11 enrolled in a high-achieving Indonesian public senior high school. The school is known for its strong academic culture and emphasis on inquiry-oriented science instruction. Students were distributed across three natural science classes (11B, 11C, and 11F). Participation was voluntary, and all data were collected with school approval and in accordance with established ethical procedures.

2.3 Instrument development

A 35-item multiple-choice test on the reaction-rate topic was initially constructed based on the structure and operational definitions of the PISA 2015 science framework. Following pilot testing and item analysis, 25 items that met the validity criterion ($r_{p\beta} \geq 0.30$) were retained as the final instrument. These items were designed to represent three domains: context (personal, local/national, and global), scientific competencies (explaining phenomena scientifically; evaluating and designing scientific inquiry; interpreting data and evidence scientifically), and scientific knowledge (content, procedural, and epistemic knowledge). The item development process adhered to the Standards for Educational and Psychological Testing [8] and established principles of science assessment [9]. In addition, contextual scenarios, graphical representations, and data tables were integrated into the items to mirror the reasoning demands characteristic of PISA science items.

2.3.1 Content validation and review

Three experts in chemistry education reviewed the items for content alignment, clarity, contextual appropriateness, and domain consistency. Revisions were made prior to field testing.

2.3.2 Pilot testing and item analysis

A small pilot cohort (non-sample) completed the draft to examine empirical validity, reliability, difficulty (p), discrimination (e.g., upper-lower 27%/point-biserial), and distractor functioning (non-functioning distractors <5% endorsement revised/removed). Retention criteria followed CTT conventions $p \approx 0.30-0.70$ as acceptable difficulty and $D \geq 0.30$ as good discrimination prior to the main administration.

Pilot testing and item analysis were conducted using a pilot implementation administered to students outside the main sample. The item analysis procedure included calculating item difficulty indices (p -values), corrected item-total correlations ($r_{p\beta}$), and distractor analyses to evaluate the plausibility and functioning of each distractor option [10,11]. Items with $r_{p\beta} \geq 0.30$ were retained, in line with widely accepted criteria for ensuring adequate internal consistency and construct discrimination [9,10]. Based on the response distribution patterns, all distractors were found to function adequately, indicating that students successfully attracted students with lower understanding while remaining less attractive to students with higher understanding.

2.3.3 Reliability

Internal consistency reliability was evaluated using Cronbach's alpha, which is an appropriate coefficient for dichotomously scored items [12]. The final instrument showed acceptable reliability for descriptive domain-level reporting.

2.4 Data Collection

The final instrument was administered during regular chemistry instructional time. Students were given 45 minutes to complete the test individually without external resources. Responses were scored dichotomously (1 = correct, 0 = incorrect).

2.5 Data Analysis

Because the purpose of this study was descriptive domain mapping, the data analysis was limited to calculating percent-correct scores for each item and domain, as well as computing descriptive statistics, including the mean, standard deviation, median, minimum, and maximum. In addition, the interquartile range (IQR) was used as a robust index of dispersion, appropriate for data that may not follow a normal distribution [13]. The results were further summarized as domain-level profiles covering context, scientific competencies, and scientific knowledge. No inferential analyses (t-tests, ANOVA, regression, or normality tests) were conducted, as the study did not aim to compare groups or establish causal effects. Moreover, percent-correct results obtained from classroom assessments were not interpreted as equivalent to PISA ability-scale scores [1–3].

3. Result and Discussion

3.1 Overall Performance on the Reaction-Rate Scientific Literacy Test

A total of 102 students completed the 25-item PISA-aligned scientific literacy assessment on the reaction-rate topic. Students' overall performance yielded a mean score of 75.29% (SD = 20.65), with a median of 84%, and a score range of 24–100%. The interquartile range (IQR) was 57–96%, indicating that the middle 50% of students performed within this interval. These values suggest that most students achieved relatively high levels of performance, with moderate variability across the sample.

3.2 Performance Across the PISA 2015 Domains

Consistent with the PISA 2015 science framework, students' scientific literacy was analysed across three domains: context, scientific competencies, and scientific knowledge. Table 1 presents the descriptive statistics for each domain.

Table 1. Descriptive statistics for scientific literacy across the three PISA domains.

Domain	Mean (%)	SD	Median	Min–Max
Context	75.17	20.63	83	25–100
Scientific competencies	76.12	22.74	83	17–100
Scientific knowledge	73.47	18.92	75	25–100

Students performed highest in the scientific competencies domain (M = 76.12%), suggesting relative strengths in interpreting data, evaluating inquiry, and applying reasoning to contextual problems. In contrast, scientific knowledge (M = 73.47%) showed slightly lower performance, indicating possible gaps in conceptual, procedural, or epistemic understanding relevant to reaction-rate phenomena.

3.3 Class-Level Descriptive Statistics

Descriptive results for each class (11B, 11C, and 11F) are presented in Table 2. While minor differences were observed across classes, these differences were not analysed inferentially, as the study did not aim to compare instructional treatments or demographic backgrounds.

Table 2. Class-level descriptive statistics.

Class	N	Mean (%)	SD	Median	IQR	Min–Max
11B	34	77.35	18.91	87	67–97	33–100
11C	34	74.71	21.80	83	56–94	25–100
11F	34	73.82	21.39	83	54–96	24–100
Overall	102	75.29	20.65	84	57–96	24–100

Class 11B exhibited a slightly higher mean and narrower IQR than Classes 11C and 11F, suggesting relatively greater score homogeneity. However, given that the study did not involve experimental manipulation or group comparisons, these descriptive differences are presented only to characterise the data, not to support inferential claims.

3.4 Domain-Level Profile

To provide a visual overview of the domain distribution, Figure 1 shows the mean percent-correct scores for each of the three PISA science domains. The pattern shows a modest but consistent tendency for students to score higher on contextual reasoning and scientific competencies than on scientific knowledge, echoing findings from prior studies that inquiry-oriented classrooms often strengthen data interpretation skills more than deep conceptual understanding.

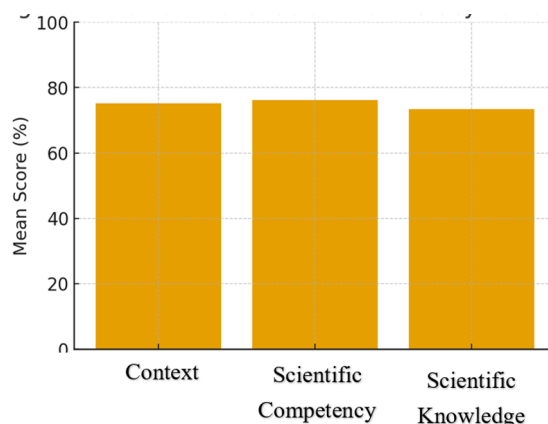


Fig. 1. Domain-level scientific literacy performance (percent-correct) for context, scientific competencies, and scientific knowledge.

4 Discussion

The results of this study indicate that students demonstrated relatively high performance on the reaction-rate scientific literacy test, with an overall mean score of 75.29%. While these findings appear substantially higher than Indonesia's performance in PISA 2018 and PISA 2022 [2,3], the difference must be interpreted with caution. As emphasised in the PISA technical framework, percent-correct scores from classroom-based assessments cannot be directly compared with PISA

ability-scale scores, which are derived from item-response modelling and calibrated across a large international pool of heterogeneous items [1]. The higher scores observed in this study, therefore, do not imply that students have achieved PISA-equivalent proficiency, but rather reflect performance on a bounded, topic-specific assessment aligned to the reaction-rate content taught in school. Similar patterns of difficulty have been observed even among first-year university students, who tend to struggle with mechanistic and epistemic aspects of chemical kinetics despite adequate procedural familiarity.

Class-level results showed only modest variation across classes, with Class 11B exhibiting slightly higher homogeneity as reflected in its narrower interquartile range. These descriptive differences likely reflect natural variability in classroom composition rather than instructional differences, as no experimental or comparative treatments were applied.

At the domain level, students performed highest on scientific competencies ($M = 76.12\%$), followed by context and scientific knowledge. This pattern aligns with evidence suggesting that inquiry-oriented instructional environments tend to strengthen students' abilities to interpret graphs, analyse data, and reason with evidence, skills central to scientific competencies [7,14]. The participating school is known for its strong academic culture and emphasis on guided inquiry, which may help explain, *in a non-causal, descriptive sense*, why students excelled in data interpretation and contextual reasoning tasks embedded in the assessment.

In contrast, the slightly lower performance in the scientific knowledge domain ($M = 73.47\%$) suggests the persistence of conceptual or epistemic gaps related to reaction-rate concepts. Research in chemistry education consistently documents misconceptions about collision theory, concentration effects, temperature, surface area, and catalysts [4–6]. Such misconceptions can reduce accuracy on items that require mechanistic explanations or the integration of procedural and epistemic knowledge, even when students can successfully interpret data patterns. The present findings thus support prior arguments that inquiry-based instruction, although valuable, may not automatically ensure deep conceptual restructuring unless explicitly scaffolded to address underlying misconceptions [7,12].

The stronger performance in the context domain indicates that students were able to recognise and reason about reaction-rate phenomena situated in personal, local, or global contexts. This is consistent with the broader goals of Indonesia's current curriculum reforms, which emphasise contextualised, competency-based learning aligned with the *Profil Pelajar Pancasila*. However, without complementary strengthening of conceptual and epistemic knowledge, contextual reasoning may remain superficial rather than mechanistic.

These findings highlight the importance of domain-level diagnostic assessment in chemistry education. By examining performance across the domains of context, scientific competencies, and scientific knowledge as conceptualised in the PISA 2015 framework, teachers can obtain a nuanced understanding of students' strengths and areas requiring deeper instructional support. In the present study, the results illustrate a pattern typical of inquiry-oriented classrooms: strong data-analytic reasoning but lingering gaps in conceptual understanding. These insights can inform instructional refinement, such as integrating more explicit conceptual scaffolding, targeted feedback, and epistemic prompts into inquiry-based reaction-rate lessons.

5 Conclusion

This study demonstrated the proficiency of Grade 11 students in a high-performing Indonesian public school in understanding reaction rates, aligning with the PISA test criteria. Overall, the students did well, especially in science and understanding the context. Students were good at understanding information, using facts to support their points, and putting what students knew to use in real life. But the fact that the students did a little worse on the science test shows that the conceptual and epistemic gaps people often talk about in reaction-rate learning are still there. You

shouldn't think of these scores as the same as PISA ability-scale scores because the two tests use very different ways to measure things. The results show that domain-level diagnostic assessment can help make teaching better instead. The domain pattern here shows that the person is good at asking questions but still doesn't fully understand some concepts. So, inquiry-based teaching could be even better if it had clear conceptual scaffolding and tailored feedback for each student. Researchers may utilize samples from multiple schools, incorporate constructed-response items, and connect assessments to broader metrics of the quality of inquiry-based learning in the future.

Acknowledgment

We thank to Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia for financial support this research. Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology Catalyst Research Scheme (Strategic Research Collaboration) Contract Number: 085/C3/DT.05.00/PL/2025, April 4th 2025.

References

1. OECD, *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving* (OECD Publishing, Paris, 2017). <https://doi.org/10.1787/9789264281820-en>.
2. OECD, *PISA 2018 Results – Country Note: Indonesia* (OECD Publishing, Paris, 2019). https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/pisa/publications/national-reports/pisa-2018/featured-country-specific-overviews/PISA2018_CN_IDN.pdf
3. OECD, *PISA 2022 Results – Country Note: Indonesia* (OECD Publishing, Paris, 2023). https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/11/pisa-2022-results-volume-i-and-ii-country-notes_2fca04b9/indonesia_0e09c072/c2e1ae0e-en.pdf
4. S. Chairam, N. Klahan, R. Coll, Exploring secondary students' understanding of chemical kinetics through inquiry-based learning activities, *Eurasia Journal of Mathematics, Science and Technology Education* **11**(5), 937–956 (2015). <https://doi.org/10.12973/eurasia.2015.1365a>
5. I. Rahayu, S. Mulyani, T. Widhiyanti, Analysis of misconceptions on the factors that affect the reaction rate, *KnE Social Sciences* **9**(13), 140–150 (2024). <https://doi.org/10.18502/kss.v9i13.15914>
6. J. Jusniar, E. Effendy, E. Budiasih, S. Sutrisno, Misconceptions in rate of reaction and their impact on misconceptions in chemical equilibrium, *European Journal of Educational Research* **9**(4), 1405–1423 (2020). <https://doi.org/10.12973/eu-jer.9.4.1405>
7. K.M. Jegstad, Inquiry-based chemistry education: a systematic review, *Studies in Science Education* **60**(2), 251–313 (2024). <https://doi.org/10.1080/03057267.2023.2248436>
8. R.L. Ebel, D.A. Frisbie, *Essentials of Educational Measurement*, 5th edn. (Prentice-Hall, Englewood Cliffs, NJ, 1991).
9. S. Hassan, R. Hod, Use of item analysis to improve the quality of single best answer multiple-choice questions in summative assessment of undergraduate medical students in Malaysia, *Education in Medicine Journal* **9**(3), 33–43 (2017). <https://doi.org/10.21315/eimj2017.9.3.4>
10. S. Testa, A. Toscano, R. Rosato, Distractor Efficiency in an Item Pool for a Statistics Classroom Exam: Assessing Its Relation With Item Cognitive Level Classified According to Bloom's Taxonomy. *Frontiers in Psychology* **9**, 1585 (2018). <https://doi.org/10.3389/fpsyg.2018.01585>

11. M. Tavakol, R. Dennick, Making sense of Cronbach's alpha, *International Journal of Medical Education* **2**, 53–55 (2011). <https://doi.org/10.5116/ijme.4dfb.8dfd>
12. A.W. Lazonder, R. Harmsen, Meta-analysis of inquiry-based learning: Effects of guidance, *Review of Educational Research* **86**(3), 681–718 (2016). <https://doi.org/10.3102/0034654315627366>
13. B. Öztürk, M. Kaya, M. Demir, Does inquiry-based learning model improve learning outcomes? A second-order meta-analysis, *Journal of Pedagogical Research* **6**(4), 201–216 (2022). <https://doi.org/10.33902/JPR.202217481>
14. M. Muntholib, S. Ibnu, S. Rahayu, F. Fajaroh, S. Kusairi, B. Kuswandi, Chemical literacy: Performance of first year chemistry students on chemical kinetics, *Indonesian Journal of Chemistry*, 20(2) , 468-482 (2020). <https://doi.org/10.22146/ijc.43651>