JPPIPA 11(12) (2025)

Jurnal Penelitian Pendidikan IPA

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

Beyond Content Validity: Comprehensive Validation of Scientific Literacy Assessment for Junior High School Teachers

Kintan Limiansih[1*], Albertus Hariwangsa Panuluh[2], Niluh Sulistyani[3]

[1] Primary Teacher Education Study Program, Faculty of Teacher Training and Education, Universitas Sanata Dharma, Yogyakarta, Indonesia.
[2] Physics Education Study Program, Faculty of Teacher Training and Education, Universitas Sanata Dharma, Yogyakarta, Indonesia.
[3] Mathematics Education Study Program, Faculty of Teacher Training and Education, Universitas Sanata Dharma, Yogyakarta, Indonesia.

Received: October 06, 2025
Revised: November 14, 2025
Accepted: December 29, 2025
Published: December 31, 2025

Corresponding Author:
Kintan Limiansih
kintan@usd.ac.id

DOI: 10.29303/jppipa.v11i12.13043

Abstract: Despite the central role of teachers in fostering students' scientific literacy, assessment instruments targeting in-service science teachers remain limited. This study aimed to develop and examine the preliminary quality of a scientific literacy instrument for junior high school science teachers, informed by OECD documents outlining the direction of the PISA 2025. A design-based research approach was employed using Tessmer's formative evaluation model, encompassing iterative stages of preliminary analysis to pilot field test. The instrument comprised nine context-rich stimuli and 25 items across multiple formats designed to elicit scientific reasoning and decision-making in real-world contexts. Expert validation produced a Mean Expert Score of 4.67, while teacher readability evaluation yielded a Mean Readability Score of 4.29, indicating strong content representation and clarity. Pilot psychometric analysis showed a balanced distribution of item difficulty (32% easy, 40% moderate, and 28% difficult), and item-level validity and discrimination indices provided diagnostic evidence for refinement. Pilot psychometric analysis indicated that the instrument's reliability remains preliminary and highlights the need for further refinement through larger-scale field testing. Overall, this study contributes a forward-looking assessment instrument that supports formative evaluation of teachers' scientific literacy and provides a robust foundation for subsequent large-scale validation aligned with emerging international assessment orientations.

Keywords: Design-based research; Junior high school; Scientific literacy; Teacher

Introduction

Scientific literacy has become a central competence in contemporary education, particularly in preparing citizens to engage with complex socioscientific issues related to public health, environmental sustainability, and technological change (Osborne & Allchin, 2024; Schenk et al., 2021; Sjöström, 2024). International frameworks conceptualize scientific literacy not merely as mastery of scientific concepts, but as the capacity to interpret scientific information, evaluate evidence, and apply scientific reasoning for informed decision-making in real-world contexts (OECD, 2023c; Schenk et al., 2021). Scientific literacy plays a strategic role in advancing the Sustainable Development Goals (SDGs), particularly those related to quality education, public health, and environmental sustainability, by equipping individuals to critically engage with scientific knowledge and contribute to sustainable societal development (Osborne & Allchin, 2024; Sjöström, 2024; United Nations, 2025)

Moving beyond mastery of subject content to also include the ability to interpret scientific information, evaluate evidence, and make appropriate decisions on socio-scientific issues (Coppi et al., 2023; OECD, 2023b; Roy et al., 2025). As an essential competency, literacy is regarded as one of the main pillars for achieving the Sustainable Development Goals (SDGs) (Mckay, 2018), particularly in ensuring inclusive and quality education. Literacy also supports progress toward other SDGs, including reducing social inequality, increasing

How to Cite:
Limiansih, K., Panuluh, A. H., & Sulistyani, N. (2025). Beyond Content Validity: Comprehensive Validation of Scientific Literacy Assessment for Junior High School Teachers. *Jurnal Penelitian Pendidikan IPA*, *11*(12), 802-813. https://doi.org/10.29303/jppipa.v11i12.13043

employment opportunities, and improving understanding of health and climate change issues, all of which contribute to sustainable development in a holistic sense (Kioupi & Voulvoulis, 2019; Sjöström, 2024).

In the PISA conceptualization, scientific literacy is placed at the core of preparing citizens who can participate responsibly in society (OECD, 2023a). It includes competencies such as explaining phenomena scientifically, designing and evaluating investigations, and interpreting data and evidence, which applied within personal, local, and global contexts (OECD, 2023c). However, scientific literacy outcomes in Indonesia remain a serious concern. In PISA 2022, Indonesian students performed far below the OECD average, with more than 60% scoring below proficiency Level 2 in science (OECD, 2023b). Findings from Indonesian studies also report low scientific literacy among junior high school students across science topics and learning settings (Hasasiyah et al., 2020; Jamaluddin et al., 2019; Yusmar & Fadilah, 2023). These results suggest that strengthening scientific literacy in lower secondary education is still an urgent need.

Teachers play a central role in shaping students' scientific literacy because they translate curriculum aims into classroom practice, guide inquiry and reasoning, and help students connect science ideas with real-world problems. Evidence from teacher education research shows that effective professional development—continuous training, mentoring, and collaboration—supports better instructional quality and student learning outcomes (Darling-Hammond et al., 2017; Sæleset & Friedrichsen, 2021). At the classroom level, Indonesian research also indicates that literacy-oriented learning and assessment practices can support students' critical thinking and science understanding (Hartina et al., 2020; Jamaluddin et al., 2019). In short, improving student scientific literacy is difficult without also strengthening teachers' competence and assessment practice. This is a crucial issue, as research has shown that teachers' scientific literacy has a significant impact on student learning outcomes (Habibi & Suparman, 2020; Yusmar & Fadilah, 2023).

Yet, strengthening teachers requires more than general encouragement; it requires clear diagnostic tools. In teacher professional development, a suitable framework and valid instruments are needed to map teachers' strengths and areas that require support (Eliyawati et al., 2023; Fe Bustamante & EMercado, 2024). However, most scientific literacy assessments and instrument development studies in Indonesia still focus on students or pre-service teachers (Hidayah & Rusilowati, 2019; Rahmadani et al., 2018; Suwono et al., 2022). Many instruments are developed for students or for specific content areas in the previous studies, for example, scientific literacy tests linked to particular topics or enrichment materials, and critical thinking instruments validated through Rasch analysis (Aryani et al., 2024; Faisal et al., 2023; Mulyana & Desnita, 2023; Widiatmo et al., 2019). Some studies also highlight literacy-related instructional approaches (e.g., PBL-STEM) that may support student outcomes (Parenta et al., 2022). These contributions are valuable, but they also show that standardized instruments targeting in-service junior high school science teachers remain limited.

Local evidence further supports the importance of this work. A survey of 62 science teachers in Yogyakarta reported that 61% of teachers claimed they understood scientific literacy (Limiansih et al., 2024). However, self-reported understanding does not automatically reflect the ability to apply scientific literacy concepts in teaching or assessment. This creates a practical need for a validated instrument that can examine teachers' scientific literacy more objectively and provide feedback that is useful for professional development (Hung & Wu, 2024).

Therefore, this study aims to develop a scientific literacy instrument for in-service junior high school science teachers and to report its content validation through layered validation stages. The development is guided by internationally recognized frameworks, particularly the PISA 2025 Science Framework (OECD, 2023c), and follows a multi-stage process that includes expert review, one-to-one interviews with practitioners, and small-scale field trials—steps commonly recommended in instrument development and formative evaluation (Coppi et al., 2023; Darman et al., 2024; Tessmer, 2005).

The instrument developed in this study was informed by the draft framework and conceptual directions outlined in OECD documents for the PISA 2025 cycle, rather than by an official or finalized assessment framework (OECD, 2023c). This forward-looking alignment was deliberately chosen because the target population of the study is in-service junior high school science teachers, for whom assessment instruments serve not only a diagnostic function but also a professional learning and anticipatory function (Sæleset & Friedrichsen, 2021; Kang et al., 2025). By engaging teachers with emerging emphases in scientific literacy, such as contextual reasoning, decision-making, and epistemic understanding, the instrument aims to provide a futuristic learning experience that introduces teachers to the evolving direction of international science assessments, while remaining grounded in the current empirical challenges highlighted by PISA 2022 (OECD, 2023a; Osborne & Allchin, 2024).

Method

This study uses a Design-Based Research (DBR) approach, which seeks to develop and validate scientific literacy instruments in real-world settings. The approach follows an iterative cycle of design, implementation, evaluation, and revision, with the goal of producing an instrument that is both theoretically sound and practically useful (McKenney & Reeves, 2025). For the development process, the Tessmer model was applied, consisting of several stages: preliminary investigation, self-evaluation, expert review and one-to-one sessions, small-group testing, and field trials (Tessmer, 2005). The overall process is illustrated in Figure 1.

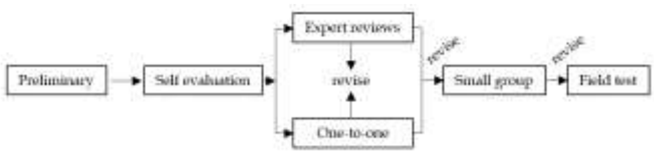

Figure 1. Stages of design and development research based on the Tessmer model

Instrument development followed the Tessmer model, consisting of four stages as follows: Stage (1), Preliminary, focused on identifying and understanding the context and objectives of the product development. At this stage, a literature review was conducted on Indonesian PISA student data and the role of teachers, highlighting the link between teacher literacy skills and student literacy outcomes. Stage (2), Self-evaluation, involved developing the initial version of the scientific literacy instrument and reviewing its alignment with the PISA 2025 Science Framework. Stage (3) included expert review of the instrument and one-to-one evaluations with junior high school science teachers to examine item readability. Stage (4) consisted of a small-group trial with a limited number of respondents. The present study was carried out up to Stage (4), focusing on construct validity and content review, and this stage also allowed the identification of fundamental issues such as item ambiguity, inappropriate difficulty levels, and technical problems before wider implementation. Stage 5 (Field test) involved administering the instrument to a broader group of respondents to further examine the preliminary psychometric characteristics of the test using classical test theory (CTT).

Participants were recruited using voluntary sampling, reflecting teachers' willingness to participate in the study. The expert involved in the review stage was selected through purposive sampling, based on demonstrated expertise and extensive experience in developing and evaluating PISA-based science literacy assessments. Participants included one expert with extensive experience in science literacy instruments, one junior high school science teacher for the one-to-one evaluation, eight science teachers for the small-group trial, and twelve science teachers for the field test.

The scientific literacy instrument was developed in alignment with the PISA 2025 Science Framework, integrating content knowledge, scientific competencies, and context. In accordance with the framework, the instrument addressed three types of scientific knowledge: content knowledge, procedural knowledge, and epistemic knowledge (OECD, 2023c). Items were embedded within a variety of real-world stimuli and situated across three science contexts: personal, local/national, and global. The application of science and technology was organized around five thematic areas: health and disease, natural resources, environmental quality (including climate change), hazards, and the frontiers of science and technology. All five areas were represented in the instrument, with contextual variation across items to reflect the breadth of scientific literacy as defined in the PISA framework.

Experts reviewed each item using a 5-point scale (1 = very poor to 5 = excellent) across four aspects: content, construction, language, and norms. A Mean Expert Score (MES) was calculated for each item using the formula:

$$MES = \left(\frac{\sum score\ of\ all\ aspects}{number\ of\ aspects}\right) \quad (1)$$

Items with MES < 3.5 were major revised. In the one-to-one phase, a teacher completed the instrument and evaluated its readability using a 7-aspect checklist (e.g., clarity of instructions, wording, and illustrations) on a 5-point scale. A Mean Readability Score (MRS) was calculated as:

$$MRS = \left(\frac{\sum score\ Aaccros\ 7\ aspects}{7}\right) \quad (2)$$

Items with MRS < 4.0 were revised for clarity and accessibility. Additional qualitative feedback was used to guide revisions.

The small-group trial tested the revised instrument with eight teachers to examine its practicality and clarity under classroom-like conditions. Participants provided written feedback that was used to refine the final version. Data analysis combined quantitative and qualitative approaches. Quantitative analysis applied MES and MRS to evaluate item quality and readability, while qualitative analysis involved coding open-ended responses using Miles and Huberman's model (data reduction, display, and conclusion drawing). This combination ensured both technical validity and practical usability of the instrument before its application in the second year of research.

The instrument was administered to twelve junior high school science teachers in a small-group trial. Responses were scored dichotomously (1 = correct, 0 = incorrect). In addition to qualitative feedback, classical test theory (CTT) analyses were conducted to examine the preliminary psychometric characteristics of the items. Item-level analyses included: validity (given the limited sample size, the results were interpreted cautiously and used to identify items requiring revision rather than to establish definitive construct validity); reliability, estimated using the Kuder–Richardson Formula 20 (KR-20); difficulty index (p); and discrimination index (D).

Result and Discussion

The instrument was developed with reference to the science literacy framework outlined in OECD documents informing the PISA 2025 cycle (OECD, 2023), particularly regarding the integration of content knowledge, scientific competencies, and contextual application. This forward-looking alignment was intended to support teachers' professional readiness by familiarizing them with emerging directions of international science assessments and the evolving expectations of scientific literacy beyond current classroom practices.

The instrument consists of two main components: a stimulus and a set of questions. This structure follows the guidelines of the Center for Educational Assessment (Pusmendik, 2019), which state that one principle in developing instruments for higher-order thinking skills is the inclusion of a stimulus (Pusmendik, 2019). A stimulus functions as a medium to encourage thinking and may take the form of text, images, scenarios, tables, graphs, discourse, dialogue, videos, or problems. According to Pusmendik (2019), the stimulus should be educational, broaden knowledge and insight, provide a positive message for behavioral improvement, and inspire those working on the instrument. An example is presented in Figure 2, which uses a procedural infographic to explain the mechanism of how vaccines work. This stimulus not only educates about the process of antibody formation in the body but also conveys a positive message regarding public awareness of the importance of vaccination.

For example, the vaccine stimulus addressed content knowledge related to immune response, procedural knowledge through interpretation of explanatory diagrams, and epistemic knowledge by requiring justification of scientific claims about vaccine effectiveness. The stimulus was situated in a personal context and targeted competencies related to explaining phenomena scientifically and evaluating scientific information for decision-making.

Figure 2. Example of a stimulus illustrating the mechanism of vaccine action, designed to be educational and to convey a positive message about the importance of vaccination

The developed stimuli were followed by accompanying questions. Each stimulus was linked to 3–4 questions, presented in a range of formats including multiple-choice, complex multiple-choice, categorical (true–false), and matching or sequencing. The distribution of question formats is presented in Table 1. Both the stimuli and the questions were constructed with reference to the PISA 2025 science framework, encompassing content knowledge, scientific competencies, and context, while the overall composition of questions was limited to the mapping shown in Table 2.

Table 1. Composition of Question Formats

Question Format	Number of Items
Multiple choice	7
Complex multiple choice: multiple answers	5
Complex multiple choice: categorical	4
Matching	2
Sequencing	3
Short answer	2
Essay	2
Total	25 items

Based on Table 1, the most frequently developed question type is multiple choice. This format was prioritized because Indonesian teachers are generally more familiar with multiple-choice items, which are widely used, relatively easy to construct, and commonly encountered in examinations (Alam & Kamela, 2022; Rezeki & Lubis, 2022). Such familiarity also helps minimize technical errors in question construction. Nevertheless, other question formats were included to ensure that the level of cognitive demand could be aligned with the scientific literacy competencies that served as the basis for question development.

The framework was operationalized by mapping items across content knowledge, procedural knowledge, and epistemic knowledge, as well as personal, local/national, and global contexts. Questions were developed to cover all five areas, with contextual variation across the items.

Table 2. Composition of Content Knowledge, Scientific Competencies, and Context in the Developed Scientific Literacy Instrument

Competencies	Explain phenomena scientifically			Construct and evaluate designs for scientific enquiry and interpret scientific data and evidence critically			Research, evaluate, and use scientific information for decision making and action		
Context	Personal	Local /national	Global	Personal	Local /national	Global	Personal	Local/national	Global
Health & Disease	Vaccination		Food security	Vaccination		Food security	Vaccination		Food security
Natural Resources	Types of personal food and energy	Renewable energy sources		Types of personal food and energy				Renewable energy sources	
Environmental Impacts & Climate Change					Waste management			Waste management	
Hazards	Lifestyle/behavior risks	Rapid changes (earthquakes, severe weather)		Lifestyle/behavior risks	Rapid changes (earthquakes, severe weather)		Lifestyle/ behavior risks	Rapid changes (earthquakes, severe weather)	
Contemporary Scientific and Technological Advances and Challenges		Use of new technology			Use of new technology			Use of new technology	

A crucial component in developing questions is the use of stimuli. In this study, nine stimuli were created as the basis for 25 questions. The topics included how vaccines work in the body, global food security maps, LiDAR technology for identifying regional conditions, frozen rice innovations for diets, renewable energy use in Indonesia, waste production in Yogyakarta, air quality in Yogyakarta, sedentary lifestyle, and natural disasters in Indonesia. The selection and design of these stimuli were guided by the content knowledge, scientific competencies, and contexts outlined in the PISA 2025 science framework.

One example is the stimulus on vaccines. The idea was adapted from the PISA 2025 framework (OECD, 2023c), which addresses issues in the personal context such as health, accidents, nutrition, vaccination, material consumption, food types, and personal energy. The focus of the vaccine stimulus was on explaining the mechanism by which vaccines build antibodies in the body, rather than on national or global vaccination data. From this stimulus, several questions were developed for instance, explaining how vaccines help prevent influenza, identifying appropriate statements to weigh risks and benefits of vaccination, and justifying research that evaluates vaccine effectiveness in protecting against viral exposure. These questions were designed to reach the level of decision-making based on scientific arguments, reflecting current challenges such as public hesitancy about vaccination.

Another stimulus addresses food security in a global context (OECD, 2023c). In the PISA framework, global issues include pandemics, food security, healthy lifestyles, renewable and non-renewable energy sources, natural systems, and related themes. The stimulus was presented as a map showing populations in different countries that cannot afford healthy food (Department of Economic and Social Affairs, United Nations). From this stimulus, questions were developed to assess skills in summarizing data and identifying strategies for policy-making in nutrition intervention programs aimed at reducing non-communicable diseases. The map was also combined with climate data from several countries to show how environmental conditions influence food security. Other stimuli were similarly developed based on the PISA framework while also considering current societal issues.

Instrument Content Validity

The quality of the instrument items in representing content knowledge, scientific competencies, and the context of scientific literacy was first evaluated through expert review. In addition to content, the experts also assessed construction, language, and norms. The results of this evaluation are presented in Figure 3.

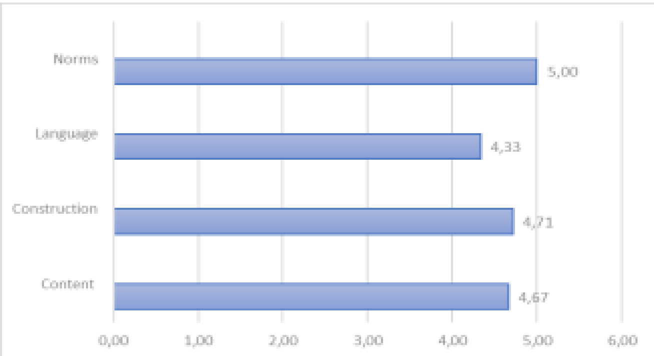

Figure 3. Expert validation scores of the instrument items

Based on the diagram in Figure 3, the instrument achieved the maximum score of 5.00 in the norms aspect, indicating very good quality. The developed stimuli and questions did not include content related to ethnicity, religion, race, and intergroup relations, nor did they contain elements that could advantage or disadvantage particular groups. They also avoided references to politics, pornography, commercial product or agency promotions, violence, or other content with potentially negative effects.

For the language aspect, the instrument obtained a score of 4.33, categorized as good. The stimuli and questions were written in accordance with Indonesian language rules and were generally communicative. However, some foreign terms still appeared in the stimulus questions; these require attention and should be translated, especially when they involve general terms (see Figure 4).

The construction aspect received a score of 4.71, also in the good category. Items were formulated clearly, aligned with the content, context, and competencies being measured, and did not provide unintended clues to the correct answer. Illustrations and other supporting elements were clear and functional, and item responses were not dependent on one another. Revisions were made to strengthen the independence of questions so that no item provided hints for another. The clarity and functionality of illustrations also need further attention, particularly regarding size, color, and the meaningfulness of visual information. Color gradation was noted as a sensitive issue; for example, images with similar colors (Figure 4) have the potential to cause subjective interpretations by respondents.

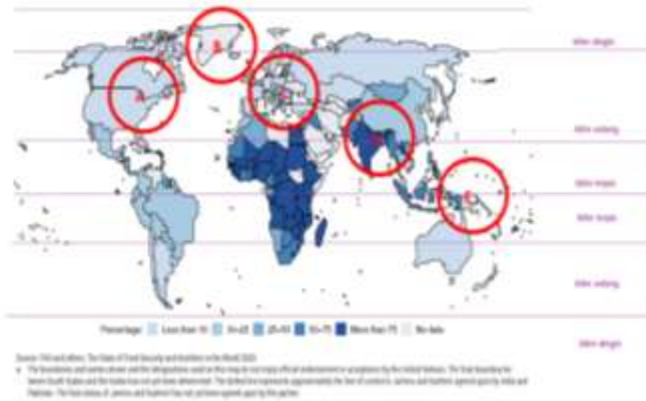

Figure 4. Example of an infographic stimulus that presents general information in a foreign language and uses color gradations, which may cause ambiguity in interpretation

In terms of content, the score obtained was 4.67. The quality of this aspect relates to the accuracy, currency, and clarity of the stimulus, as well as the ability of the questions to measure literacy skills and provide correct or workable answers. The question format can strongly influence this. For instance, multiple-choice multiple-answer items or categorical true/false items can become subjective if the statements involve habits, general knowledge, or facts that are widely recognized as true but are not included in the stimulus. Such cases may cause debate about which answer is correct. In the developed instrument, one true/false item contained a statement that could reasonably be categorized as "true" but could also be seen as "false" since it was not explicitly supported by the stimulus. This item therefore required modification.

Another important factor is the effectiveness of the scientific concepts presented in the stimuli and questions. As previously noted, the stimulus is intended to provide insight, so it must include concepts that are accurate and up to date. For example, one question described LiDAR as "a system that performs remote measurement and sensing using light emissions that can be operated by drones." This description was not scientifically precise and needed revision.

When the four aspects were combined, the average expert validation score was calculated as shown in Equation (3):

$$MES = \left(\frac{70}{15}\right) = 4.67 \quad (3)$$

Overall, the expert content validation score was 4.67, which falls into the “good” category. The experts concluded that the questions were suitable for use with only minor revisions. Alongside the expert assessment of the instrument, a readability review was conducted by junior high school science teachers through one to one focus group discussions. The readability assessment covered several aspects: clarity of instructions, clarity of the purpose of the stimulus and questions, the quality of sentences in the stimulus and questions (which needed to be communicative and free of ambiguity), the comprehensibility of illustrations such as pictures, tables, and diagrams, and the likelihood that the questions could be answered. The readability score given by the teachers was calculated using formula (2) and produced the following result:

$$MRS = \left(\frac{30}{7}\right) = 4.29 \quad (4)$$

The score in Equation (4) shows that the questions had a good level of readability. The teachers suggested adding punctuation guidelines to the instructions to help in completing the questions, since a variety of formats were included in the instrument. They also recommended adding information to the stimulus to indicate the number of questions that could be answered from each stimulus. This suggestion was reasonable, as the practice of linking one stimulus to 3–4 questions is more common in language tests. The teachers’ final suggestion was similar to the experts’: improving the clarity of images in the stimuli (see Figure 2). The combined input from experts and teachers was used as a guide for refining the questions.

Instrument Quality Based on Teacher Responses in a Small Test

After being evaluated by experts and junior high school science teachers in the one-on-one stage, the scientific literacy instrument was revised and then tested in the next phase, a small-group trial. Eight teachers completed the 25 questions and provided feedback. The percentage of time spent on the questions is presented in Figure 5.

This instrument was designed to be completed within a maximum of 80 minutes, equivalent to two periods of junior high school science lessons in Indonesia. As shown in Figure 5, most teachers were able to finish the instrument in less than 80 minutes. This indicates that the instrument is time-efficient and can be implemented as planned. It also suggests that teachers were able to use the time to think through the questions with minimal technical difficulties. As previously noted, the small-group participants were heterogeneous, involving teachers from various districts in the Yogyakarta region. The majority (75%) completed the instrument within the target time, indicating that it is feasible for use by teachers across diverse contexts in the Special Region of Yogyakarta. Nevertheless, several items were considered highly difficult and required longer time to complete (8 out of 25 items). The composition of difficult and time-consuming items, based on teacher responses, is presented in Table 3.

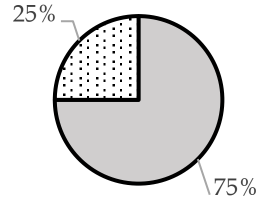

Figure 5. Percentage of time spent by junior high school science teachers in completing the scientific literacy instrument

Table 3 shows that the content most frequently identified as difficult was health and disease, both in personal and global contexts. At the personal level, the stimulus used concerned how vaccines work. Vaccination is familiar to Indonesians because it is mandatory for the public, especially for toddlers. However, the microscopic mechanism of how vaccines function in the human body was not easily understood by respondents.

Preliminary psychometric analyses were conducted using classical test theory based on a field test with 12 junior high school science teachers. Item validity, examined through corrected item–total correlations, showed that 3 of 25 items (12%) exhibited relatively strong associations with the total test score ($r = 0.67$–0.72), while the remaining items showed correlation coefficients close to zero or varying in magnitude. Two items (8%) could not be estimated because all respondents selected incorrect options, resulting in zero variance. These patterns provide early evidence of how teachers engaged with context-based scientific literacy items and reflect the cognitive and contextual demands involved in integrating scientific knowledge, competencies, and real-world situations.

Instrument reliability was estimated using the KR-20 coefficient, yielding a value of −0.73 in the present pilot sample and interpreted as preliminary psychometric evidence. Analysis of item difficulty indicated that 8 items (32%) were classified as easy ($p > 0.70$), 10 items (40%) as moderate ($0.30 \le p \le 0.70$), and 7 items (28%) as difficult ($p < 0.30$), including two items with $p = 0$. Item discrimination analysis showed that 1 item (4%) had a discrimination index of $D \ge 0.40$, 21 items (84%) had discrimination indices between $D = 0.00$ and 0.39, and 3 items (12%) yielded negative discrimination indices, indicating reversed response patterns between higher- and lower-scoring respondents.

Table 3. Items Considered Difficult and Time-Consuming by Teachers

Number	Content	Question Format	Competency	Context	Stimulus
1	Health & Disease	Multiple Choice, Multiple Answer	C1	Personal	How Vaccine Work
2*	Health & Disease	Categorical	C3	Personal	How Vaccine Work
5	Health & Disease	Multiple Choice, Multiple Answer	C1	Global	The state of food security and nutrition
9	Scientific and Technological Advances & Challenges	Multiple Choice	C3	Local	Drone Technology for Identifying Regional Conditions
14	Natural Resources	Sequencing	C3	Local	Utilization of Renewable Energy in Indonesia
15*	Enviromental Impacts	Categorical	C2	Local	Waste Data in Yogyakarta
21	Hazards	Multiple Choice	C2	Personal	Sedentary Lifestyle
22	Hazards	Multiple Choice	C3	Personal	Sedentary Lifestyle

*Items with a high frequency of being identified as difficult and requiring more time to complete.

Overall, the psychometric findings are positioned as exploratory and formative, consistent with pilot testing in design-based instrument development. Rather than supporting final claims of measurement quality, the results offer empirically grounded insights into item behaviour, contextual load, and response variability among science teachers. These findings inform targeted item revision and provide a reflective foundation for subsequent large-scale testing aimed at strengthening the robustness of scientific literacy assessments for teachers.

Teachers reported difficulties across a range of question types, including multiple-choice, multiple-answer, categorical, and sequencing items. This indicates that, although multiple-choice questions are widely used and familiar to teachers, they are not necessarily easy to complete. In terms of scientific literacy competencies, 50% of the difficult items involved the ability to research, evaluate, and use scientific information for decision-making and action. With respect to context, teachers struggled with items in personal, local, and global contexts, although the personal context appeared most challenging. This suggests that, despite its direct connection to everyday life, personal issues relating to individuals are not yet fully understood comprehensively.

Of the eight items considered difficult and time-consuming, two (items 2 and 15) were reported with the highest frequency (Figure 6). Although they differed in content, competency, context, and stimulus, they shared the same question format: categorical. Category questions consist of several statements for which respondents must decide whether each statement is true or false. Viewed through Bloom's taxonomy, this format requires higher-order thinking skills, particularly evaluation.

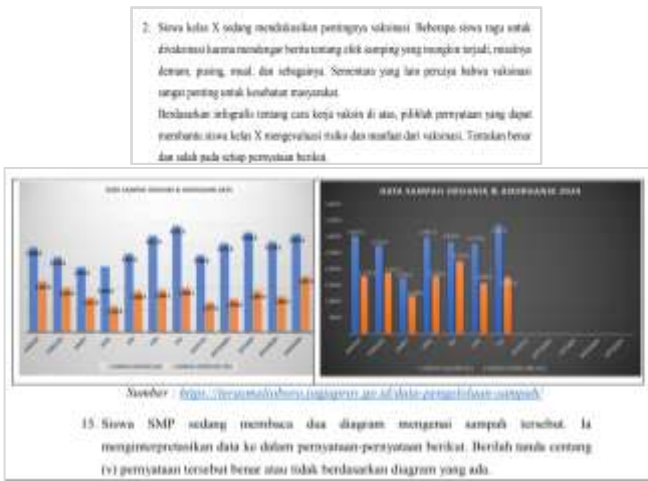

Figure 6. Examples of two scientific literacy items (on vaccination and waste management) that teachers reported as difficult and requiring more time to complete

The goal of science education is not simply to prepare a small group of students to become future scientists. As Wieman (2007) notes, what is needed is a society with strong scientific literacy to meet the global challenges humanity now faces. These challenges can only be understood—and potentially addressed—by using science as a foundation for informed decision-making (Roy et al., 2025; Tasquier et al., 2022). For this reason, sustainability-oriented science has become a priority. The present study emphasizes that developing questions with a sustainability orientation is essential, as this provides direction for planning teacher competency development to meet such challenges. Integrating science and society within the framework of sustainable development is therefore fundamental (Sjöström, 2024; Wieman, 2007), ensuring that science is seen as more than a set of concepts, facts, or laws.

One characteristic of PISA science questions is their holistic content, which integrates different fields within science (biology, physics, chemistry) as well as the relationship between science, technology, and society. The context of science learning should not be limited to classroom situations but should also connect to the knowledge and experiences already familiar to 15-year-olds. It must be relevant to their interests and directly linked to everyday life (OECD, 2023b). For this reason, the issues highlighted include health and disease, natural resources, environmental quality (including environmental impacts and climate change), hazards, and the scientific and technological advances (including contemporary advances and challenges).

Developing literacy questions therefore requires careful attention to integrated content so that both teachers and students are trained to solve problems in complex situations. Integrated science bridges the natural and life sciences, supporting deeper and more comprehensive understanding of concepts and their applications (Kelp et al., 2023). However, one major barrier to integration is the division of expertise among junior high school teachers. Interviews with teachers involved in the small-group test revealed that their specialized backgrounds in biology, physics, or chemistry limited their ability to develop holistic, cross-disciplinary competencies. Teachers trained in a single discipline often find it difficult to implement interdisciplinary learning, mainly due to a lack of experience and insufficient guidance on how to integrate different fields (Tonnetti & Lentillon-Kaestner, 2023; Tripp & Shortlidge, 2019).

Sustainability-related science issues also call for the use of integrative concepts. However, an analysis of environmental education textbooks shows that the materials commonly used by teachers and students remain largely content-oriented (Eliyawati et al., 2022). This suggests a need for teachers to strengthen their ability to update their understanding of contextual issues and to independently construct integrative networks of scientific concepts. Findings from the focus group discussions indicated that, after working with the scientific literacy instrument, teachers were able to identify topics and issues relevant to everyday life. They also reported broader insights into scientific concepts and recognized that science is not limited to personal concerns but is connected to global challenges.

The instrument component measuring the ability to research, evaluate, and use scientific information for decision-making and action was still found to be difficult by teachers. This suggests that teachers themselves continue to face difficulties in integrating theory with practice (Roy et al., 2025). Another challenge is the use of digital tools: limited skills in operating complex software contribute to suboptimal digital literacy, which in turn constrains teachers' ability to assess accurate scientific information for decision-making (Rasimin et al., 2024). For this reason, instrument development needs to maintain a balanced proportion across competencies to ensure that scientific literacy is assessed holistically.

This study also has limitations in the instrument testing stage. The developed instrument has not yet been fully examined for reliability and validity through statistical analysis, as it has only reached the small-group testing phase. This stage focused on collecting qualitative feedback to obtain initial insights into item clarity and usability. Field trials that enable more comprehensive quantitative testing therefore remain a priority for future research.

Conclusion

This study successfully developed a scientific literacy instrument for junior high school teachers. The construction of the instrument was designed in accordance with the PISA framework, taking into account the aspects of context, content, and scientific literacy competencies. The literacy questions used contexts close to real life and required reasoning skills to solve, differing from routine questions that are procedural and not contextual. Validation results showed an MES score of 4.67 in the good category, while the readability aspect obtained an MRS score of 4.29, also in the good category. Teacher responses further confirmed that this instrument can measure teachers' scientific literacy, accommodate reasoning, and inspire the introduction of topics related to current real-world problems. Preliminary psychometric analysis from the pilot test ($N = 12$) showed that 12% of items demonstrated strong item–total correlations ($r = 0.67$–0.72), the KR-20 coefficient was −0.73, item difficulty was distributed across easy (32%), moderate (40%), and difficult (28%) categories, and item discrimination indices ranged from negative values to $D \geq 0.40$, providing formative evidence for item revision. The novelty of this study lies in its focus on in-service junior high school teachers, unlike most previous studies which concentrated on students or pre-service teacher candidates. The initial trial through a small-group test indicated that the instrument can provide qualitative insights useful for further refinement, although the study has not yet reached the stage of field testing or predictive and longitudinal reliability and validity analysis. Thus, this research contributes theoretically by enriching the understanding of scientific literacy from the perspective of teachers, while also providing a practical assessment instrument with potential to support the professional development of science teachers.

Acknowledgments
The authors would like to express their sincere gratitude to validator, for the valuable contribution in this research. Special thanks are also extended to the junior high school teachers in Yogyakarta who participated in the one-to-one sessions and the small group test, as well as to the Science Subject Teacher Working Group (MGMP IPA) of the Yogyakarta Special Region (DIY) for their support in data collection and constructive discussions.

Author Contributions
Kintan Limiansih: Conceptualization; methodology; formal analysis; investigation; resources; data curation; writing original draft preparation; visualization; supervision; project administration; funding acquisition. Albertus Hariwangsa Panuluh: Conceptualization; methodology; investigation; writing, review and editing; translation; instrument development. Niluh Sulistyani: Validation; writing, review and editing; strengthening the introduction; instrument development. All authors have read and approved the published version of the manuscript.

Funding
This research was funded by the Institute for Research and Community Service (LPPM), Universitas Sanata Dharma, under the internal research grant scheme with contract No: 016 Penel./LPPM-USD/Iy2025.

Conflicts of Interest
The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. All procedures in this study were conducted in accordance with the ethical standards of educational research. The researchers obtained permission from the Provincial Education Office of Yogyakarta as well as from the five respective District Education Offices. Participation of respondents was voluntary, informed consent was ensured, opportunities for clarification were provided, and the confidentiality of all respondent data was strictly maintained.

References

Aryani, S., Novia, H., & Setiawan, A. (2024). Development and Analysis Validation Instrument Critical Thinking Heat Concept (CTHT) Using Rash Model. *Jurnal Penelitian Pendidikan IPA*, *10*(7), 3731–3738. https://doi.org/10.29303/jppipa.v10i7.7727

Coppi, M., Fialho, I., & Cid, M. (2023). Developing a Scientific Literacy Assessment Instrument for Portuguese 3rd Cycle Students. *Education Sciences*, *13*(9). https://doi.org/10.3390/educsci13090941

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Retrieved from https://learningpolicyinstitute.org/product/teacher-prof-dev.

Darman, D. R., Suhandi, A., Kaniawati, I., Samsudin, A., & Wibowo, F. C. (2024). Development and Validation of Scientific Inquiry Literacy Instrument (SILI) Using Rasch Measurement Model. *Education Sciences*, *14*(3). https://doi.org/10.3390/educsci14030322

Eliyawati, E., Widodo, A., Kaniawati, I., & Fujii, H. (2022). Education For Sustainable Development (ESD) In Students' Textbooks. In *Journal of Engineering Science and Technology Special Issue on ICMScE2022*. Retrieved from https://www.ipcc.ch/report/ar4/wg1/

Eliyawati, Widodo, A., Kaniawati, I., & Fujii, H. (2023). The Development and Validation of an Instrument for Assessing Science Teacher Competency to Teach ESD. *Sustainability (Switzerland)*, *15*(4). https://doi.org/10.3390/su15043276

Faisal, S., Rusilowati, A., & Susilaningsih, E. (2023). Science Literacy Assessment Instrument for Additives and Addictive Substances: Development, Validation, and Rasch Model Analysis. *Jurnal Penelitian Pendidikan IPA*, *9*(10), 7826–7836. https://doi.org/10.29303/jppipa.v9i10.4376

Fe Bustamante, M. O., & EMercado, R. (2024). Professional Development And Scientific Literacy In Science Curriculum. *Sci-Int. (Lahore)*, *36*(6), 417–422. Retrieved from https://sci-int.com/pdf/638715227225281132.pdf

Habibi, H., & Suparman, S. (2020). Literasi Matematika dalam Menyambut PISA 2021 Berdasarkan Kecakapan Abad 21. *JKPM (Jurnal Kajian Pendidikan Matematika)*, *6*(1), 57-64. Retrieved from http://journal.lppmunindra.ac.id/index.php/jkpm/

Hartina, L., Rosidin, U., & Suyatna, A. (2020). Pengaruh Penerapan Instrumen Performance Assessment pada Pembelajaran IPA Berbasis Laboratorium Real terhadap Hasil Belajar Siswa. *Jurnal Penelitian Pendidikan IPA*, *6*(1), 25–31. https://doi.org/10.29303/jppipa.v6i1.299

Hasasiyah, S. H., Hutomo, B. A., Subali, B., & Marwoto, P. (2020). Analisis Kemampuan Literasi Sains Siswa SMP pada Materi Sirkulasi Darah. *Jurnal Penelitian Pendidikan IPA*, *6*(1), 5–9. https://doi.org/10.29303/jppipa.v6i1.193

Hidayah, N., Rusilowati, A., & Masturi, M. (2019). Analisis profil kemampuan literasi sains siswa

SMP/MTs di Kabupaten Pati. *Phenomenon: Jurnal Pendidikan MIPA, 9*(1), 36-47. https://doi.org/10.21580/phen.2019.9.1.3601

Hung, C. S., & Wu, H. K. (2024). High school science teachers' assessment literacy for inquiry-based science instruction. *International Journal of Science Education*, *46*(7), 621–642. https://doi.org/10.1080/09500693.2023.2251657

Hurian Kamela, R. S. A. (2022). Public Teacher Recruitment And Selection In Indonesia: A Legal Perspective. *Scientia Business Law Review (SBLR)*, *1*(1), 37–43. https://doi.org/10.56282/sblr.v1i1.51

Jamaluddin, J., Jufri, A. W., Ramdani, A., & Azizah, A. (2019). Profil Literasi Sains Dan Keterampilan Berpikir Kritis Pendidik IPA SMP. *Jurnal Penelitian Pendidikan IPA*, *5*(1). https://doi.org/10.29303/jppipa.v5i1.185

Kelp, N. C., McCartney, M., Sarvary, M. A., Shaffer, J. F., & Wolyniak, M. J. (2023). Developing Science Literacy in Students and Society: Theory, Research, and Practice. *Journal of Microbiology & Biology Education*, *24*(2). https://doi.org/10.1128/jmbe.00058-23

Kioupi, V., & Voulvoulis, N. (2019). Education for sustainable development: A systemic framework for connecting the SDGs to educational outcomes. *Sustainability (Switzerland)*, *11*(21). https://doi.org/10.3390/su11216104

Limiansih, K., Sulistyani, N., & Melissa, M. M. (2024). Persepsi Guru SMP terhadap Literasi Sains dan Implikasinya pada Pembelajaran Sains di Sekolah. *Jurnal Pendidikan Mipa*, *14*(3), 786–796. https://doi.org/10.37630/jpm.v14i3.1858

Mckay, V. (2018). Literacy, lifelong learning and sustainable development. *Australian Journal of Adult Learning*, *58*(3), 390–425. Retrieved from https://search.informit.org/doi/abs/10.3316/informit.069488054638414

McKenney, S., & Reeves, T. C. (2025). Educational design research for relevant & robust scholarship. *Journal of Computing in Higher Education*, *37*(2), 614–638. https://doi.org/10.1007/s12528-025-09456-2

Mulyana, V., & Desnita, D. (2023). Empirical Validity and Reliability of the Scientific Literacy Assessment Instrument Based on the Tornado Physics Enrichment Book. *Jurnal Penelitian Pendidikan IPA*, *9*(5), 3961–3967. https://doi.org/10.29303/jppipa.v9i5.3290

OECD. (2023a). *Education at a Glance 2023* (Education at a Glance). OECD Publishing. https://doi.org/10.1787/e13bef63-en

OECD. (2023b). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD. https://doi.org/10.1787/53f23881-en

OECD. (2023c). *PISA 2025 Science Framework* (Draft). Retrieved from https://pisa-framework.oecd.org/science-2025/

Osborne, J., & Allchin, D. (2024). Science literacy in the twenty-first century: informed trust and the competent outsider. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2024.2331980

Parenta, Y., Masykuri, M., & Saputro, S. (2022). Literature Study: Application of PBL-STEM on Simple Machine Topic to Improve Critical Thinking Skills. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 674–680. https://doi.org/10.29303/jppipa.v8i2.1181

Rahmadani, Y., Fitakurahmah, N., Fungky, N., Prihatin, R., Majid, Q., & Prayitno, B. A. (2018). Profil Keterampilan Literasi Sains Siswa di Salah Satu Sekolah Swasta di Karanganyar. *Jurnal Pendidikan Biologi*, *7*(3), 183–190. https://doi.org/10.24114/jpb.v7i3.10123

Rasimin, Semma, A. B., Zakiyuddin, Ali, M., & Helmy, M. I. (2024). Multi-dimensional challenges in the Indonesian social science information technology-based learning: A systematic literature review. *Heliyon*, *10*(7). https://doi.org/10.1016/j.heliyon.2024.e28706

Roy, G., Sikder, S., & Danaia, L. (2025). Adopting scientific literacy in early years from empirical studies on formal education: a systematic review of the literature. In *International Journal of STEM Education* (Vol. 12, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40594-025-00547-1

Sæleset, J., & Friedrichsen, P. (2021). Pre-service Science Teachers' Pedagogical Content Knowledge Integration of Students' Understanding in Science and Instructional Strategies. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(5), 1–18. https://doi.org/10.29333/ejmste/10859

Schenk, L., Hamza, K., Arvanitis, L., Lundegård, I., Wojcik, A., & Haglund, K. (2021). Socioscientific Issues in Science Education: An opportunity to Incorporate Education about Risk and Risk Analysis? *Risk Analysis*, *41*(12), 2209–2219. https://doi.org/10.1111/risa.13737

Sjöström, J. (2024). Vision III of scientific literacy and science education: an alternative vision for science education emphasising the ethico-socio-political and relational-existential. In *Studies in Science Education*. Routledge. https://doi.org/10.1080/03057267.2024.2405229

Rezeki, S. K., & Lubis, F. (2022). *Analisis Soal Buatan Guru Bahasa Indonesia Berdasarkan Taksonomi The Structure Of The Observed Learning Outcome (Solo) Siswa Kelas XI SMA Negeri 12 Medan Tahun*

Pembelajaran 2020/2021 (Vol. 11, Issue 1). Retrieved from http://ejournal.uika-bogor.ac.id/index.php/TEK

Suwono, H., Maulidia, L., Saefi, M., Kusairi, S., & Yuenyong, C. (2022). The Development and Validation of an Instrument of Prospective Science Teachers' Perceptions of Scientific Literacy. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(1). https://doi.org/10.29333/EJMSTE/11505

Tasquier, G., Knain, E., & Jornet, A. (2022). Scientific Literacies for Change Making: Equipping the Young to Tackle Current Societal Challenges. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.689329

Tessmer, M. (2005). *Planning and Conducting Formative Evaluations: Improving the Quality of Education and Training*. Routledge.

Tonnetti, B., & Lentillon-Kaestner, V. (2023). Teaching interdisciplinarity in secondary school: A systematic review. In *Cogent Education* (Vol. 10, Issue 1). Taylor and Francis Ltd. https://doi.org/10.1080/2331186X.2023.2216038

Tripp, B., & Shortlidge, E. E. (2019). A framework to guide undergraduate education in interdisciplinary science. *CBE Life Sciences Education*, *18*(2). https://doi.org/10.1187/cbe.18-11-0226

United Nations. (2025). *The Sustainable Development Goals Report*. Retrieved from https://unstats.un.org/sdgs

Widiatmo, T., Jufri, A. W., & Jamaluddin, J. (2019). Analysis of Validation of Instruments to Measure Student's Critical Thinking Ability and Science Literation. *Jurnal Penelitian Pendidikan IPA*, *5*(2), 212–218. https://doi.org/10.29303/jppipa.v5i2.272

Wieman, C. (2007). Why Not Try a Scientific Approach to Science Education. *The Magazine of Higher Learning*, 9–15. https://doi.org/10.3200/CHNG.39.5.9-15

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