Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram
https://e-journal.undikma.ac.id/index.php/prismasains/index
e-mail: prismasains.pkpsm@gmail.com

January 2026. Vol. 14, No. 1
p-ISSN: 2338-4530
e-ISSN: 2540-7899
pp. 1-33

Developing and Profiling the Scientific Literacy Test Items on Prospective Science Teachers: A Focus on Health and Disease in Socioscientific Issues

[1*]Winata Tegar Saputra, [1]Lilit Rusyati, [1]Nuryani Y. Rustaman, [2]Indra Fardhani, [3]Ismail Fikri Natadiwidjaya
[1]Department of Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Department of Science Education, Universitas Negeri Malang, Malang, Indonesia
[3]Department of Science Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

*Corresponding Author e-mail: winatategar08@upi.edu

Received: June 2025; Revised: October 2025; Published: January 2026

Abstract

This study aimed to develop scientific literacy instruments aligned with the 2025 framework, focusing on socioscientific issues (SSI) related to health and disease, and apply them to prospective science teachers (PSTs). Despite the updated framework, many Indonesian scholars still use outdated models. Using an exploratory sequential mixed-methods design and the ADDIE model, this research involved 194 PSTs, three assessment experts, and two content experts. Textbook analysis revealed that competency 1 was most dominant, while competency 3 was the least addressed in both books and classroom practice. The instrument development process began with 30 items and underwent multiple stages, including expert validation, readability checks, pilot testing, and statistical analyses of reliability, item difficulty, and discrimination. Ultimately, 23 items were validated with high reliability (0.92) and a balanced difficulty range that cover the cognitive skills such as content, procedural, and epistemic knowledge. A profiling analysis using these items showed that most PSTs demonstrated moderate scientific literacy, with third-year students outperforming first- and second-year students. Overall, the instrument proved effective, as it was systematically developed based on textbook analysis and teacher interviews to reflect current scientific literacy demands, particularly in preparation for PISA 2025. This instrument also expects the PSTs in the future to have the ability to teach their students about the skills of the content, procedural, and epistemic knowledge.

Keywords: Development; Health and diseases; Instrument; Scientific literacy; Socioscientific issues

How to Cite: Saputra, W. T., Rusyati, L., Rustaman, N. Y., Fardhani, I., & Natadiwijaya, I. F. (2026). Developing and Profiling the Scientific Literacy Test Items on Prospective Science Teachers: A Focus on Health and Disease in Socioscientific Issues. *Prisma Sains : Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *14*(1), 1–33. https://doi.org/10.33394/j-ps.v14i1.16435

https://doi.org/10.33394/j-ps.v14i1.16435

INTRODUCTION

There are reasons that the scores of PISA focused on scientific literacy are still low. Yusmar and Fadilah (2023) argue the reason for the low scientific literacy score is due to Teachers provide limited opportunities for students to practice scientific literacy tasks, prioritizing content mastery instead. They infrequently encourage concept development, rely on teacher-centered approaches, and often lack sufficient knowledge of scientific literacy. Additionally, the fast-paced teaching aimed at meeting curriculum goals frequently causes misconceptions about science concepts, reducing learning to rote memorization. Another reason, is the students do not learn science by contextual, so the students could not correlate the knowledge to the phenomena that happen around them (Latif et al., 2022). Also, the main cause of the low scientific literacy level among Indonesian students is most of the assessments are focused on the cognitive aspect and lack of high cognitive level (Murti & Sunarti, 2021)

The problem that faced in Indonesia regarding scientific literacy assessment is teachers' knowledge of scientific literacy is still limited, and they have not yet utilized scientific literacy test instruments (Astuti et al., 2014). Therefore, many researchers research developing scientific literacy test items with some lacknesses. The development of the test items about scientific literacy that emphasize explaining the scientific phenomenon of the subject matter of energy was conducted by the Borg and Gall model design for students in secondary school. Nevertheless, part of the instrument test is developed from UMPTN and part of it is developed from PISA 2015 (Adawiyah & Wisudawati, 2017). The instrument test that was developed shows the weaknesses, especially the differences between UMPTN that was intended for last year's high school students, and PISA that was intended for 15-year-old students who average on last year in secondary school. Another development instrument regarding scientific literacy test items was developed based on the local wisdom that contains the three competencies based on the previous PISA scientific literacy test by the OECD (Murti & Sunarti, 2021), but the context of the developing scientific literacy test does not fit into one of the contexts in PISA. Another research that used local wisdom as the main context was conducted by Maulida and Sunarti (2022) that lacks the context that does not fit into one of the contexts in PISA. Also, the sample that was intended is not suitable, with the PISA requirement that this research is intended for high school students. Based on the previous research that has been reviewed, none of the research, especially Indonesian research developed or tested the students of scientific literacy based on the PISA 2025 framework with the latest competency.

The differences PISA 2025, 2015, and 2006 frameworks are about the competencies. In PISA 2015 framework, the competencies focused on the (1) explain phenomena scientfically, by identifying, presenting, and assessing explanations for various natural and technological occurrences, (2) evaluate and design scientific enquiry, by describing experiments, evaluating their methods, and suggesting scientific approaches to answer questions, (3) Interpret data and evidence scientifically, by examining information, evaluating arguments and claims, and drawing suitable scientific conclusions (OECD, 2016). In the PISA 2006 framework, there are three competencies: (1) identifying scientific issues, by identifying problems that can be explored through scientific investigation, selecting important keywords to effectively find scientific information, understand and highlight the essential elements of a scientific investigation, (2) explaining phenomena scientifically, apply scientific knowledge to specific situations, describe, interpret, and predict scientific phenomena and their changes, choose accurate explanations, descriptions, and predictions related to scientific concepts, (3) using scientific evidence, by alyze scientific data to draw, communicate, and support conclusions, identify the assumptions, evidence, and logical reasoning that underpin conclusions, reflect on how science and technological advances impact society (OECD, 2007).

The development of scientific literacy test items was conducted by Chasanah and Colleagues (2022) using the 4D (define, design, develop, and disseminate) model, but this research discards the dissemination phase. Furthermore, the sample of this research is intended for elementary schools that are not suitable for PISA conditions that should be intended for 15-year-old students. Also, the context is not given in this research. Another research about developing scientific literacy test items by 4D, discards the dissemination phase conducted by Martinah and Colleagues (2022) in context of environmental pollution. However, the sample that are chosen by the researcher are intended for secondary school students, and not all of the secondary school students fulfill the requirements of the PISA test. The mistaken samples are collected by Rosidah and Sunarti (2017) who choose the high school students as the sample for developing scientific literacy test items. Also, the topics that were chosen are heat that is not suitable to the context of PISA scientific literacy. While, the research by Putri (2020) focused the developing scientific literacy test items by ADDIE model in the topics of diversity of living creatures. Some weaknesses shows about the sample that are choose are intented to first years of secondary school that not suitable to PISA requirements and the PISA competencies are not consider in this research. Also, the research about the development of the intrument based on PISA 2025 framework are still lacks.

To fulfill the gap from previous research, this research focused on the development of the scientific literacy test items based on PISA 2025 framework in health and disease topics to prospective science teachers that measure the unidimensional construct that measure the ability of the scientific literacy among PSTs. Therefore, the research questions consist of the following:

1. How do the students' science textbooks support scientific literacy, students' experiences about their learning and readiness, and teachers' guidance?
2. How do the content validity of the developed scientific literacy test items measure up based on experts' judgment and PSTs' readability?
3. How do the test items analysis of the developing test items by the Pilot Study on the limited and wide sample?
4. How does the profiling of PSTs scientificc literacy by using developed test items?

METHOD

Research Design

This research developed the scientific literacy test items related to PISA 2025, which is the science topic as the domain test. Also, this research determined the ability of the PSTs to meet each competency based on PISA 2025 rules. Therefore, this research used an exploratory sequential mixed methods design (Creswell & Creswell, 2023), that integrate with an ADDIE (analyze, design, develop, implement, and evaluate) approach (Branch, 2009).

Research Subject

This research assesses Prospective Science Teachers (PSTs) who are expected to be professionals in Indonesian schools from three universities and at various level of the PSTs. The sampling technique in this research used convenience sampling, which is a group of people available for this research (Fraenkel et al., 2017). The research subject consists of three universities, the detailed in Table 1

Table 1. Research Subject

University	Level	Number	Total	Percentages (%)
University X	First year	30	62	31.96
	Second year	19		
	Third year	13		
University Y	First year	0	38	19.59
	Second year	38		
	Third year	0		
University Z	First year	0	94	48.45
	Second year	94		
	Third year	0		

Data Collection

The development of the scientific literacy test items used the ADDIE (analyze, design, develop, implement, and evaluate) approach and was integrated with the exploratory sequential mixed methods design. The "analyze" stage started by analyzing the students' main and additional science textbooks and teachers' guidance books about health and disease in Kurikulum Merdeka. The analysis textbooks focused on health and disease topics, which contained the aspects of scientific literacy and socioscientific issues in content, context, and assessment. Then, three teachers were interviewed to clarify the students' experiences and readiness about scientific literacy, socioscientific issues, and health and disease. These results are the sources for the qualitative data.

In the “design” stage, the researchers construct 30 test items based on the PISA 2025 focus on the health and disease context in socioscientific issues. Next stage is the sources of the quantitative data, in the “develop” stage, the test items were conducted to assess content validity about 30 test items by three expertize in assessment and two experts in content. To find out about the invalid test items in the content validity if the Aiken validity test is low. The Aiken validity test resulted in one test item being invalid, therefore, the number became 29 test items became valid and need revisions based on the experts’ suggestions. After that, the readability test was conducted after the test items were revised by five PSTs. The next stage is the “implement” stage to reveal the validity, reliability, difficulty, and discriminatory index. In the construct validity, the researchers need to observe three categories of the value of outfit MNSQ values range between 0.5 and 1.5, the outfit ZSTD values range from -2.0 to 2.0 and the item-total correlation values fall between 0.4 and 0.85. The test items are considered valid or infit as the category requires at least 2 of 3 those categories (Sumintono & Widhiarso, 2015). If the test item do not meet those criteria, then it consider as invalid test items The pilot test was conducted on a limited sample of 30 PSTs in the first year using 29 test items, resulting in 24 valid test items. The second pilot test used 24 test items for 60 PSTs in the second and third years, resulting in 23 valid test items. Both of the pilot test results were analyzed for the “evaluate” stage, consisting of validity, reliability, difficulty, and discriminatory index. Those stages are the sources of the quantitative data. The detailed of the participants in each stage shown in Table 2.

Table 2. The Participants in each Stage of the Test

Stage of the test	Level	Number	Total
First pilot test	First year	30	30
	Second year	0	
	Third year	0	
Second Pilot Test	First year	0	60
	Second year	47	
	Third year	13	
Profiling test	First year	30	194
	Second year	151	
	Third year	13	

The profiling test was conducted to assess 194 PSTs integrated students' reulst from the first and second pilot test, and adding about 104 students for profiling, using 23 test items as a result of the development of the test items. However, the analysis also added the PSTs' answers in a first and second pilot test, analyzing 23 valid test items. therefore, the number of respondents analyzed is 194 PSTs. The analysis of the profiling test is about the analysis in descriptive statistics of their achievement regarding the scientific literacy based on the test items that have been developed and analyzed for their achievement regarding the competency, knowledge, SSI elements, and theme. Also, the comparison in each group also conducted to compare the achievement between the first year and second year, the first year and third year, and the second year and third year independent sample t-test, and also the comparison in achievement of the competency, knowledge, SSI elements, and theme. The comparison of the PSTs’ years conducted in one university. The detail is shown in Table 3.

Table 3. Participants of the Comparison Profiling Analysis for University X

PSTs level	Number	Percentage (%)
First year	30	48.39
Second year	19	30.65
Third year	13	20.97
Total	62	100

University X was chosen because the PSTs level consisted of three levels, whereas University Y and X that just have only one level of PSTs. Also, the comparison among the university was not conducted because imbalance of participants and levels that might influence the results. For example, if the researcher conducts a comparison of three universities in the second year, the numbers do not balance. To simplify the explanation of the research stage, the pattern of the research is shown in Figure 1.

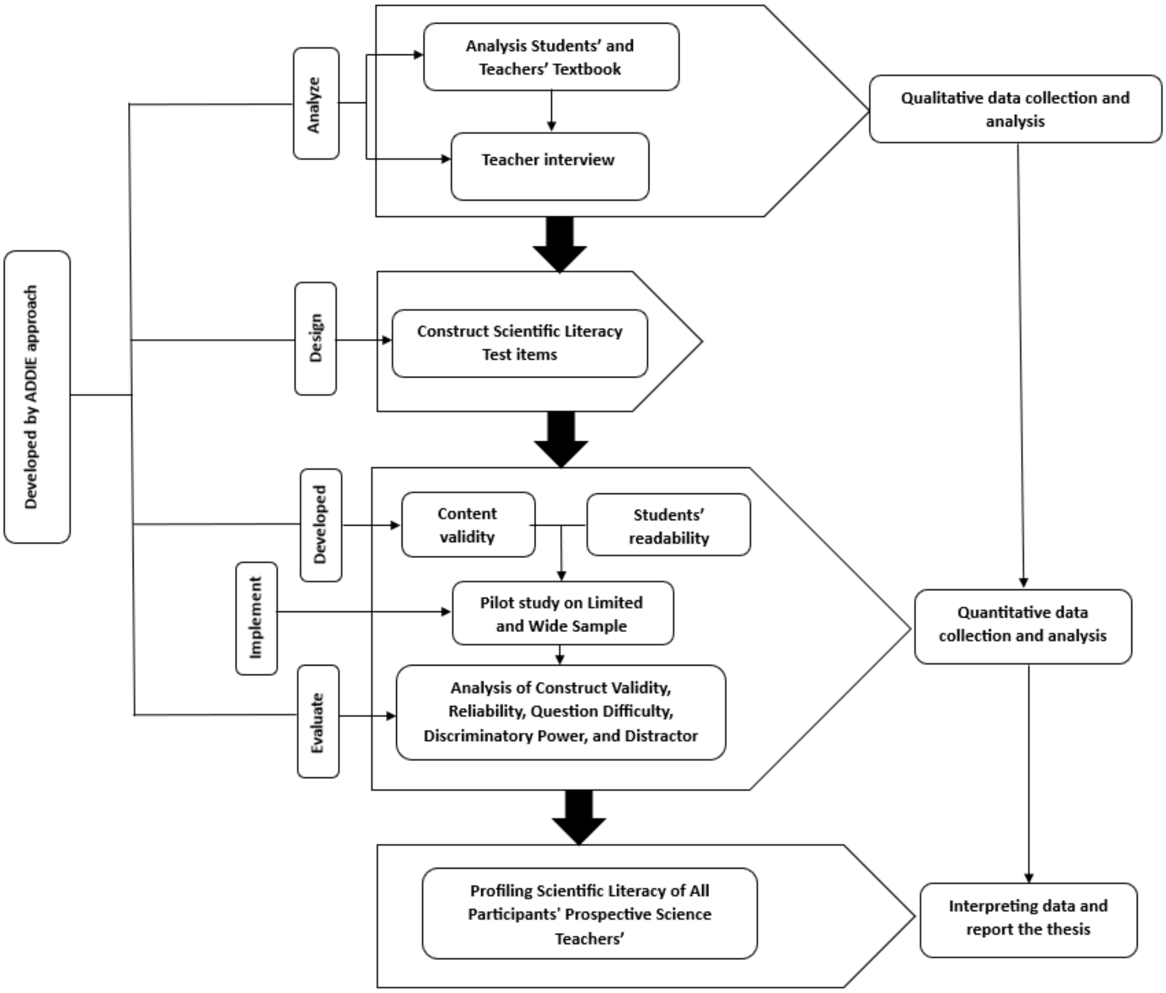

Figure 1. The Pattern of The Research

RESULTS AND DISCUSSION

The Condition of the Students' Science Textbooks in Supporting Scientific Literacy and Students' Experiences about Learning and Readiness and Teachers' Guidance

This analysis of the students' science textbooks and students' experiences about their learning and readiness is included in the "analyze" phase in the ADDIE model. This research analyzes the condition of the students' science textbooks and whether they fulfil the scientific literacy and SSI aspects and became a foundation to construct and develop scientific literacy test items. Then this sub-chapter also has the phase of "design" to fulfil the weaknesses of the students' science textbooks.

The analysis of scientific literacy competencies in the main science textbooks demonstrates distinct patterns across different chapters and sub-chapters. In the chapter Introduction to Cells, was exists in the main textbooks in eighth grade, but does not exist in the additional textbooks. The existence of this chapter relates to the health and disease related to the assessment part discusses about the stem cells in the humans. In this assessment, competency 1 requires students to explain and describe about the stem cells. In competency 2, students demanded to draw a diagram to people who support this idea. Lastly, in competency 3, students demanded to conduct research about the people who support and reject this idea. In this activity also correlate with SSI, it because students need to correlate their research to factor such as health, economy, society, culture, ethics, and environment. However, due to this chapter relate to health and disease in the final assignment, therefore, the existance of the competency and SSI in content and context is not exists.

The chapter of Structure and Function of a Living Organism has four sub-chapters on main textbooks and additional science. Competency 1 presents about explanation of the content in all sub-chapter that the presentation of those two books are similar. In the content, those books present by the explanation that occurs in the human body. The existence of competency 1 also exists in the context, that demanded students to describe their understanding in the students' activity related to the chapter. Lastly, competency 1 was exists in the final assessment to train their remembering and their explanation such as the calculation about calories on human body or the physical activity that influence the demand of the calories. Competency 2 in main textbooks requires students to calculate about the way to interpret the nutritional information of the product that exists in the content, that similar existance in the additional textbooks. while the context that about the student activity also similar about the interpret the nutritional information of the food they have choosen by themselves. In the assessment, students calculate the total of the body water that gone because of the hot weather and heavy training from the shown data. In the additional textbooks of this chapter was strong on competency 1, but lack of competency 2 and 3, and the SSI that the amount less than in the main textbooks.

In the Growth and Development chapter, the amount of competency 1 exists in the content and assessment in both the main and additional textbooks. Competency 1 in this chapter request students to explaining the phenomenon of the zygote that develops become fetus, while in the assessment request students to find out the definition and the factor that influence stunting. In the competency 2 on this chapter was exist in the assessment that request students to draw a graph of the age and the average body height and they need to interpret that. In the competency 3, the students require to suggest the best meals for reduce the stunting with the nutrition fact that similar to the additional textbooks. While the SSI aspect emphasize on the main textbooks that emphasize on the solution of the meals for reduce stunting by consider about the culture, environment, economy, and society that in the additional textbooks were not request students about this.

The chapter Human Coordination, Reproduction, and Homeostasis presents a more balanced distribution about the competencies. Both of the main and additional textbooks present almost similar. Competecy 1 present on the content that explain about the structure and function of the reproduction system on the human and explain about the connection of the body to the brain. In the context, competency 1 requests students to explain how the sound sources could affect the hearing, while in the assessment, competency 1 asked students to explain about the structure and function of the female reproduction, and the explanation of the cranium of a newborn baby is not strong. Competency 2 in the content explains about the graph of the menstruation of females with the hormon, ovarium cycle, and uterus cycle. In the context, the studenst also asked to interpreting the menstruation cycle with some cases. In the assessment, students need to find out by the inquiry strategy to find out the reasons of the smoking can narrow the artery and cause the ciruclatory system being in problem. However, this chapter do not present about the competency 3 on both of the main and additional textbooks. In SSI, it present on the main textbooks on the assessment that asked students to make project about the addresses ethical, cultural, and economic aspects of contraception, encouraging informed decision-making and awareness of social inequalities. The integration of these elements enables students to critically analyze the role of science in societal contexts, linking personal health choices with broader social and economic factors.

Notably, the chapter Environmental Issues, which discusses the health and disease topics were exists in both of the textbooks about environmental health. Competency 1 is present in content that explains about the phenomenon of environmental health in Indonesia, especially about the COVID-19 pandemic and other environmental issues that correlate with human health on human such as forest fire, safe drinking water, and other. Competecny 1 do not present about the context and assessment in this chapter. Competency 2 is present in content that help students to interpret the graph about the percenrtage of households with access to safe drinking water of the provinces in Indonesia. In the context, competency 2 and 3 asked students to assess the quality of their drinking water from various sources such as borehole, public tap, piped water, and others. This competency 2 is not exists on the assessment section. In SSI, the content in this chapter explained the problems of diseases caused by the environment in Indonesia, such as ISPA and diarrhea. Also, the government rules about safe drinking water physically, microbiologically, chemically, and radioactively. In the context, The SSI integrated the students' activity to explores the water sources to identify the quality of the water for their daily such as drinking. However, additional textbooks do not show any SSI.

The review of scientific literacy competencies within additional science textbooks reveals a slightly different distribution compared to the main textbooks. In the chapter Structure and Function of a Living Organism, particularly across the food and digestive system, circulatory system, respiratory system, and excretory system sub-chapters, a strong emphasis is placed on recognizing scientific issues and explaining scientific phenomena, especially in sub-chapters related to the food and digestive system. However, the interpretation of scientific evidence and the integration of socioscientific issues are minimal, with limited instances observed, primarily confined to final assignments. In the Growth and Development chapter, a similar pattern persists: identifying scientific issues is prioritized, while competencies in explaining phenomena and using evidence are less prominent, and socioscientific considerations are largely absent. In the chapter Human Coordination, Reproduction, and Homeostasis, the coordination system and reproductive system sub-chapters demonstrate some engagement with recognizing scientific issues, though again with minimal focus on evidence interpretation and socioscientific contexts. The Environmental Issues chapter, focusing on environmental health in Indonesia, provides more balanced competency coverage, incorporating elements of scientific explanation and evidence use, along with a modest integration of socioscientific issues. Overall, identifying scientific issues remains the most dominant competency across the additional textbooks, with a frequency of (26), followed by explaining phenomena (11) and using scientific evidence (4). Socioscientific issues appear sparingly, totaling only 5 occurrences across the chapters. This analysis suggests that, although scientific issues are well-represented, there is a clear need to strengthen the emphasis on scientific reasoning, evidence interpretation, and socioscientific connections to better support students' scientific literacy development.

The triangulation of the science textbooks analysis presents the integration of scientific literacy and SSI in a biology curriculum, categorized into chapters, sub-chapters, and three key dimensions: content, context, and assessment. The findings indicate that scientific literacy is consistently incorporated across all sub-chapters, with content being fully covered and assessed in most cases. However, the integration of SSI is less consistent, particularly in terms of contextual understanding and assessment. In the chapter on "Structure and Function of a Living Organism," for example, scientific literacy is well-represented across the food and digestive system, circulatory system, respiratory system, and excretory system, with all three dimensions of content, context, and assessment being addressed. In contrast, while SSI is introduced in these topics, its contextual relevance is often lacking and rarely included in assessments. A similar trend is observed in the "Human Coordination, Reproduction, and Homeostasis" chapter, where scientific literacy is fully integrated. However, SSI is only partially covered, particularly in the reproductive system and homeostasis sub-chapters. Furthermore, in the chapter on "Growth and Development," the scientific literacy component is thoroughly covered, while SSI lacks both content and contextual discussions, which may limit students' ability to relate biological concepts to real-world issues. A more balanced approach is observed in "Environmental Issues," where both scientific literacy and SSI are incorporated, though the assessment of SSI remains limited. Overall, the findings suggest that while scientific literacy is systematically embedded in the curriculum, the inclusion of SSI remains inconsistent.

Another result of the triangulation in main science textbooks shows that the competency to “Explain Phenomena Scientifically” is the most common in health and disease topics in Indonesian students' textbooks. Followed by “Construct and Evaluate Designs for Scientific Inquiry and Interpret Scientific Data and Evidence Critically,” which is the second highest. While “Research, Evaluate, and Use Scientific Information for Decision-Making and Action” has the lowest status. This indicates that the three scientific literacies are not equally balanced in improving students’ achievement of scientific literacy. Additionally, the aspect of SSI in the health and disease topics does not appear in all sub-chapters. Most of the SSI aspects relate to the issues from Indonesia's circumstances.

The similar finding of textbooks triangulation analysis in the additional science textbook has similar results regarding the comparison of three competencies and SSI. The competency of “Explain Phenomena Scientifically” is the most common in this textbook, and the amount of this competency in the additional science textbook is more than in the main textbook. Therefore, the addition science textbook intends to support the additional content of the chapters. Another competency such as “Construct and Evaluate Designs for Scientific Inquiry and Interpret Scientific Data and Evidence Critically,” and “Research, Evaluate, and Use Scientific Information for Decision-Making and Action”, has fewer than the main textbook. Due to the additional science textbooks' focus on the content of the chapter, then the SSI aspect shows a smaller amount than in the main textbook. However, the additional textbook support the activity of the kurikulum merdeka demand such as AKM. The pattern of the amount of the scientific literacy competencies and socioscientific issues is similar between the main science textbooks and the additional science textbooks. The “Explain Phenomena Scientifically” competency is the most exist, followed by “Construct and Evaluate Designs for Scientific Inquiry and Interpret Scientific Data and Evidence Critically” as the second highest competency, thirdly, socioscientific issues as third highest, and lastly the “Research, Evaluate, and Use Scientific Information for Decision-Making and Action” competency is the fewest.

The results are in contrast to the analysis of the scientific literacy aspects in elementary schools in the previous Indonesia curriculum (kurikulum 2013), which shows a lack of scientific literacy aspects, such as aspects of science as a way of thinking. About 5.8%, which is the highest indicator, establishes cause-and-effect relationships, followed by showing how science operates with inductive and deductive reasoning. Regarding aspects of the interaction of science, technology, and society, about 1.2%, with the highest indicator discussing social issues related to science or technology, followed by describing the usefulness of science and technology for society (Nurfaidah, 2017). A similar finding about analysis of the textbooks in buku pendamping TEMATIK IPA shows the aspects of science as a way of thinking; the highest indicator shows is establishing a cause-and-effect relationship, followed by the fact with the proof. In SSI aspects, scientific literacy, technology, and society in discussing social issues related to science or technology is the highest, followed by describing the usefulness of science and technology for society (Dewi & Putra, 2022). However, the textbook analysis results in this research are contrasted in the scientific literacy aspects, but the SSI aspects discussed the social issues that occur and are related to the topics' material.

Students' learning experiences were merely aligned with the textbooks, but it is still about the way teachers teach. Therefore, this research conducted an interview with three science teachers who responded about scientific literacy, SSI, and health and disease. Teachers implement scientific literacy in different ways, with some administering PISA scientific literacy test items to assess students' competencies and prepare them for standardized tests like ANBK (Asesmen Nasional Berbasis Komputer). Others incorporate literacy-based assessments in summative evaluations rather than using standardized PISA questions. A major challenge faced in fostering scientific literacy is students' difficulty in interpreting graphs and tables. Some teachers integrate data interpretation activities into their instruction, while others allocate minimal time to such tasks due to students' struggles with fundamental concepts. Scientific literacy is not only about content knowledge but also about reasoning and decision-making. While some teachers focus on delivering fundamental concepts, others actively engage students in reasoning activities through case studies, structured questioning, and guided analysis.

Rohim and Colleagues (2021) argue that ANBK plays the role of assessing students' literacy and numeracy at elementary and secondary levels to measure what they have learned and to analyse the learning experiences in the schools. Numerical literacy skills are related to the ability to apply basic knowledge, principles, and mathematical processes to real-life problems, such as understanding issues presented in tables or diagrams, trade, and others. Numerical literacy differs from mathematical competence, as the distinction lies in how concepts and knowledge are utilised. Many schools have poor preparation regarding this assessment because the schools have short periods to prepare students for facing the assessment, and the questions in the national assessment do not match the material that students have learned in the classes (Sholikhah & Purwani, 2023). These reasons are the foundation for science teachers to implement scientific literacy from PISA to encourage students to improve their literacy ability. The preparation of facing ANBK is expected for students to be ready. This preparation also shows the positive results proven by previous research by Aini and Pramasdyahsari (2023) that shows good results about the ability literacy and numeracy, and their ability to operate the computer for conducting the national assessment

The integration of SSI into science instruction is widely acknowledged as beneficial for enhancing scientific literacy. Teachers utilize real-life examples to improve comprehension, integrating discussions on topics such as stunting, pollution, global warming, and addictive substances. These SSI contexts help students connect scientific knowledge to real-world phenomena and encourage critical thinking. By engaging with SSI, students develop the ability to analyze scientific concepts in meaningful contexts and make informed decisions. The incorporation of SSI into the curriculum is seen as a foundational strategy for strengthening scientific literacy and fostering analytical skills. Implementing SSI learning by Indonesian science teachers shows some challenges, Nida and Colleagues (2020) argue that the absence of essential student competencies causes the lack of SSI learning implementation, insufficient teacher expertise, limitations in the official curriculum content, inadequate facilities, and insufficient time for lesson planning and execution. However, those teachers can recommend a suitable chapter that has the potential for implementing SSI learning. Also, teachers' perspective and attitudes about SSI show positive results (Faisal & Martin, 2022). The previous research is aligned with the results of the interview that show a positive attitude towards SSI, and even the teachers used SSI to simplify the explanation in instruction. Additionally, most teachers lack knowledge about SSI because teacher training is still rare, but Indonesia found that SSI started to train prospective teachers in lesson study (Genisa et al., 2020).

The Content Validity of The Developed Scientific Literacy Test Items Measure Up Based on Experts' Judgment and Students' Readability

This research question was included in the phase of "develop" in the ADDIE model. The suggestions of the content validity consist of accept with value 2, accept with revisions with value 1, and reject with value 0. In all of the experts' suggestions, the results are analyzed by Aiken validity. Also, the readability of the test items after content validity revisions was conducted for prospective science teachers. The aiken validity of 5 experts, consisting of 3 experts in assessment and 2 experts in content, was analyzed. The results of the content validity by Aiken Test are shown in Table 4.

Table 4. Result of Content Validity by Aiken Test

Question	Assessment experts' results	Content experts' results
1	Invalid	Medium valid
2	Medium valid	Valid
3	Valid	Valid
4	Valid	Valid
5	Medium valid	Valid
6	Valid	Valid
7	Valid	Medium valid
8	Valid	Medium valid
9	Medium valid	Medium valid
10	Valid	Medium valid
11	Valid	Medium valid
12	Valid	Medium valid
13	Valid	Medium valid
14	Valid	Medium valid
15	Valid	Valid
16	Valid	Medium valid
17	Medium valid	Medium valid
18	Medium valid	Medium valid
19	Valid	Medium valid
20	Medium valid	Medium valid
21	Valid	Medium valid
22	Valid	Medium valid
23	Medium valid	Valid
24	Valid	Valid
25	Valid	Medium valid
26	Valid	Valid
27	Medium valid	Valid
28	Valid	Medium valid
29	Valid	Medium valid
30	Valid	Medium valid

The valid score based on assessment experts' decision has 21 test items valid and 8 test items medium valid, while the invalid score has 1 test item. Based on the content experts' decision, ten test items are valid and 20 test items are medium valid. After the instrument was conducted the content validity, the instrument was revised based on the experts' suggestions. The results show that test item number 1 is not valid by the Aiken test based on the assessment experts' decision, which resulted in the test items becoming 29. Therefore, the blueprint such as competency of "construct and evaluate designs for scientific enquiry and interpret scientific data and evidence critically", procedural knowledge, complex multiple-choice, SSI element of "integrate implicit and/or explicit ethical components that require some degree of moral reasoning", and coronavirus pandemic theme are deleted for one number. while the recapitulation of the content validity such as the amount of valid, medium valid and invalid between the assessment expert and content expert that shown in Table 5.

Table 5. The Content Validity Recapitulation

Category	Assessment experts	Content experts
Valid	21	10
Medium valid	8	20
Invalid	1	0

Then, the readability test was conducted by 5 prospective science teachers. The readability test aims to evaluate the test items to understand, ensure clarity, and reduce the ineffective sentences in each test item for the intended participants. The results of the 5 prospective science teachers about the test items are about “readable” or “revise” which are shown in Table 6.

Table 6. Result of Prospective Science Teachers (PST) Readability Test

Test item	PST 1	PST 2	PST 3	PST 4	PST 5
Q2	Readable	Readable	Readable	Revise	Readable
Q3	Readable	Revise	Readable	Revise	Revise
Q4	Revise	Readable	Readable	Revise	Readable
Q5	Readable	Readable	Readable	Revise	Readable
Q6	Revise	Revise	Revise	Readable	Readable
Q7	Readable	Readable	Readable	Revise	Revise
Q8	Readable	Revise	Readable	Readable	Readable
Q9	Readable	Readable	Revise	Readable	Revise
Q10	Revise	Readable	Readable	Readable	Revise
Q11	Revise	Readable	Readable	Readable	Revise
Q12	Revise	Readable	Readable	Revise	Revise
Q13	Readable	Readable	Readable	Readable	Readable
Q14	Revise	Readable	Readable	Readable	Readable
Q15	Readable	Readable	Readable	Revise	Readable
Q16	Readable	Readable	Readable	Readable	Readable
Q17	Readable	Readable	Readable	Readable	Readable
Q18	Revise	Readable	Readable	Readable	Readable
Q19	Revise	Revise	Readable	Revise	Readable
Q20	Revise	Revise	Readable	Revise	Revise
Q21	Revise	Readable	Readable	Readable	Readable
Q22	Revise	Readable	Readable	Readable	Readable
Q23	Readable	Readable	Readable	Revise	Revise
Q24	Readable	Readable	Readable	Readable	Readable
Q25	Readable	Readable	Readable	Readable	Revise
Q26	Revise	Readable	Readable	Readable	Readable
Q27	Readable	Readable	Readable	Readable	Readable
Q28	Readable	Readable	Readable	Readable	Readable
Q29	Readable	Readable	Readable	Readable	Readable
Q30	Revise	Readable	Readable	Readable	Readable

The readability test results for prospective Science Teachers (PSTs) reveal a clear divide between consistently well-received questions and those requiring revision. Items such as Q13, Q16, Q17, Q24, Q27, Q28, and Q29 were unanimously rated as "Readable" by all five PSTs, confirming their clarity and effectiveness. However, other questions particularly Q6, Q12, Q19, and Q20 were frequently flagged for revision, suggesting issues with wording, complexity, or ambiguity. Mixed responses for items like Q3, Q9, Q23, and Q25 further highlight the need for minor adjustments to ensure uniform readability. These discrepancies may stem from differences in PSTs' interpretations of the questions or varying thresholds for clarity.

Notably, PST 1 and PST 4 were more critical in their evaluations, marking a higher number of items as "Revise" compared to PST 2, PST 3, and PST 5, who generally found most questions readable. This divergence underscores the importance of reviewing contested items to align them with broader readability standards. To enhance the test’s reliability, priority should be given to revising questions with the most "Revise" ratings, especially those with strong disagreement. Additionally, gathering targeted feedback from PSTs on problematic items could help refine ambiguous phrasing and ensure the test is both fair and accessible to all respondents

Other research conducted the content validity of the test items by Aiken validity conducted by three experts, which are two chemistry lecturers and one chemistry teacher. The researchers constructed 15 test items and the results show 12 test items categorized as valid and 3 test items categorized as medium valid (Wahyuni & Yusmaita, 2020). The Aiken validity test was also conducted to develop the test items for literacy and numeracy as the subscale of the National Assessment in Indonesia for elementary school students. The results of the Aiken validity test show a result of 0.88, meaning the test items are valid by analyzing the aspect of Clarity, accuracy, relevance, and validity (Kusuma & Nurmawanti, 2023). The validation of the Aiken test validity conducted by 8 science teachers by a focus group discussion to develop the instrument of literacy and numeracy for the Asesmen Kompetensi Minimum (AKM). The validity of the develop instrument results from 20 test items become 18 test items. No information regarding the category of the validity index of the Aiken test of the research (Nabil et al., 2022). Other research conducted the scientific literacy test items, content validation, beside Aiken test validity. The content validity by two experts of content that shows an average validity of 86%, which is categorised as high validity and two experts of media that shows the average is 95%, which is categorised as high validity (Hasanah et al., 2024). Other than the development of the scientific literacy test items based on local wisdom, conducted by the content validity by two experts, judged on the content, has a score of 87.7% which categorizes as very valid by using theoretical validation. While scoring on the construction of the test, has a score of 94.01%. In the language area score as 87.7% (Murti & Sunarti, 2021). Other results that show similar content validity by face validity to expertize on the material that received an average score of 93%. The scientific literacy aspect also obtained an average percentage of 93%, while the construction aspect achieved 99%. Additionally, the grammatical aspect was rated with an average percentage of 95% (Chasanah et al., 2022).

In the readability test, this research uses five PSTs with has similar characteristics with future respondents for future steps. Twenty-nine test items were conducted using the readability test by the evaluator, which has similar characteristics to the prospective participants for this research. The results of the readability test do not affect the validity or reliability of the test items, and conducting the adjustment of the sentences in the questions or the options before the test items are assessed to the participants. The suggestions from the prospective science teachers regarding the test items are shown in Appendix. The suggestions by the PST are expected to be revised and adjusted to the test items. The evaluation of test items highlights key areas for improvement, particularly in language accuracy, clarity, and adherence to EYD. Common errors include incorrect spelling such as “mempengaruhi” instead of “memengaruhi” or influence and improper capitalization in pronouns like “Anda,” or “You” while foreign terms like “GERD” and “stunting” should be italicized. Ambiguous phrasing and redundant words, such as “masih bisa mungkin,” should be simplified for clarity, and vague expressions like “pusat pemerintahan dan akses yang mudah” require specification. Structural issues, including ineffective question transitions and punctuation errors, need refinement to enhance readability. Additionally, key terminology like "6M 1S," or “GERD”, should be clearly defined to ensure comprehension. Several questions require rewording to improve clarity, with a focus on eliminating repetition and restructuring sentences for better readability. These revisions aim to enhance the precision, coherence, and effectiveness of the test items, ensuring they meet academic writing standards and facilitate clearer understanding for respondents.

The results and discussion contain scientific research findings and discussions. Write down scientific findings obtained from the results of research that has been done but must be supported by adequate data. The scientific findings referred to here are not the results of the research data obtained. The scientific findings must be explained scientifically including: What scientific findings were obtained? Why did that happen? Why are trend variables like that? All these questions must be explained scientifically, not only descriptive if necessary supported by adequate scientific basis phenomena. In addition, it should also be explained in comparison with the results of other researchers who are almost the same topic. The results of the research and findings must be able to accommodate the research objectives in the introduction.

The Test Items Analysis of The Developing Test Items by The Pilot Study on The Limited and Wide Sample

The result of the pilot test which includes the "implement" phase of the ADDIE model, while the analysis such as construct validity, reliability, question difficulty, and discriminatory power was included in the "evaluate" phase of the ADDIE model. The construct validity and reliability were conducted twice, for a limited sample of 30 PSTs and a wide sample of 60 PSTs, using the Rasch model, Winstep analysis. The construct validity assesses PSTs for 29 test items after the test items, after the content validity, and PSTs' readability. After conducting the construct validity on the limited sample, the analysis continued into the question of difficulty, and discriminatory power. The results of the validity in the Rasch model, Winstep analysis for limited and wide samples are shown in Table 7.

Table 7. Result of Construct Validity in Limited Sample and Wide Sample

Test item	Result for Limited Sample	Result for Wide Sample
Q2	Fit/valid	Fit/valid
Q3	Fit/valid	Fit/valid
Q4	Fit/valid	Fit/valid
Q5	Fit/valid	Fit/valid
Q6	Fit/valid	Fit/valid
Q7	Fit/valid	Fit/valid
Q8	Misfit/invalid	Not analyzed
Q9	Fit/valid	Fit/valid
Q10	Misfit/invalid	Not Analyzed
Q11	Fit/valid	Fit/valid
Q12	Fit/valid	Fit/valid
Q13	Fit/valid	Fit/valid
Q14	Fit/valid	Fit/valid
Q15	Fit/valid	Fit/valid
Q16	Misfit/invalid	Not analyzed
Q17	Misfit/invalid	Not analyzed
Q18	Fit/valid	Fit/valid
Q19	Fit/valid	Fit/valid
Q20	Fit/valid	Fit/valid
Q21	Fit/valid	Fit/valid
Q22	Fit/valid	Fit/valid
Q23	Fit/valid	Fit/valid
Q24	Misfit/invalid	Not analyzed
Q25	Fit/valid	Fit/valid
Q26	Fit/valid	Fit/valid

Test item	Result for Limited Sample	Result for Wide Sample
Q27	Fit/valid	Fit/valid
Q28	Fit/valid	Fit/valid
Q29	Fit/valid	Fit/valid
Q30	Fit/valid	Misfit/invalid

From the Rasch model, Winstep analysis shows 24 test items were fit/valid test items, which means that 5 test items were expressed as not fit/invalid test items in the pilot test to a limited sample. Then, five test items decided to be not fit/invalid were not used for further analysis in the pilot test for a wide sample. The result of the analysis between the limited and wide samples is shown in Table 8.

Table 8. The Results of The Analysis between Limited and Wide Sample

Analysis	Limited sample	Wide sample
Construct validity	24 test items	23 test items
Reliability		
Person reliability	0.61	0.71
Item reliability	0.88	0.92
Competency		
Explain phenomena scientifically	10 test items	10 test items
Construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically	7 test items	7 test items
Research, evaluate, and use scientific information for decision-making and action	7 test items	6 test items
Knowledge		
Content	10 test items	10 test items
Epistemic	7 test items	6 test items
Procedural	7 test items	7 test items
Types of test items		
Simple multiple-choice	14 test items	13 test items
Complex multiple-choice	10 test items	10 test items
Themes		
Corona virus	5 test items	5 test items
Stunting	4 test items	4 test items
Smoking	4 test items	4 test items
Diabetes	5 test items	5 test items
Acute respiratory infections	6 test items	5 test items
SSI element		
Employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation	8 test items	7 test items
Integrate implicit and/or explicit ethical components that require some degree of moral reasoning	3 test items	3 test items
Emphasize the formation of virtue and character as long-range pedagogical goals	7 test items	7 test items
Utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics	6 test items	6 test items

The results of the reliability in the limited sample have two types, as this research is about the development of the test items, then it must consider the item reliability, which scored 0.88, considered as a good reliability category. While for a wide sample pilot test, it found the item reliability is higher than previous. It found that the item reliability is 0.92, which is expressed as a good reliability category similar to previous reliability. When the pilot test on a limited and wide sample has been conducted and shows that the not fit or not valid test items, then there are differences in the frequencies among the scientific literacy competencies, knowledge, themes, and types of questions between the pilot test on limited and wide sample.

The results show that the competency of "explain phenomena scientifically" is the most frequently tested item in this research. Other competencies such as "construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically" and "research, evaluate, and use scientific information for decision-making and action" have similar frequencies. However, when the second pilot test was conducted and test item 30 was not fit/ invalid, then the amount of the competency "research, evaluate, and use scientific information for decision-making and action" was reduced. Besides the competency, there is knowledge of each test item. There are content knowledge, procedural knowledge, and epistemic knowledge. The frequencies of the knowledge in the test items have similarity to the competencies. It shows that the content knowledge is the most frequent in the test items, while other knowledge such as epistemic knowledge and procedural knowledge have the same frequencies. However, it is similar to the previous case in that in the second pilot test, there is test item 30 that deleted because of invalid. Therefore, the amount of knowledge is reduced, and the epistemic knowledge is reduced by one number. The number of types of test items after conducting the pilot test on a limited sample shows that the simple multiple-choice has more items than the complex multiple-choice. However, the second pilot test shows the amount of the simple multiple-choice is reduced by one number, but the comparison between them was similar also after the test items conducted the pilot test on a wide sample that shows the simple multiple-choice has more numbers than complex multiple-choice.

The result of the pilot test on a limited sample shows that the acute respiratory infection theme is the highest number, while the second highest is the theme of corona virus and diabetes with a similar number of the test items. Lastly, the smallest number of test items is stunting and smoking with four test items for both of them. A pilot test on a wide sample shows only one test item that is not fit/invalid in the topic of acute respiratory infections. Therefore, the amount of the test items in that theme is reduce. The SSI elements show the differences between previous construction before the pilot test. It shows that the highest element of SSI is "employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation", while the second highest SSI element is "emphasize the formation of virtue and character as long-range pedagogical goals" that has similar amount with the element "utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics". Lastly, the smallest amount of SSI elements "integrate implicit and/or explicit ethical components that require some degree of moral reasoning". After the results of the pilot test have been revealed, the analysis of the difficult test items is conducted to find which test items are the most difficult for a limited sample. The analysis used the Rasch model, winstep analysis and the figure of variable (wright) map. The result of the difficulty test items from the variable (wright) map for the pilot test on limited sample and wide sample is shown in Figure 2.

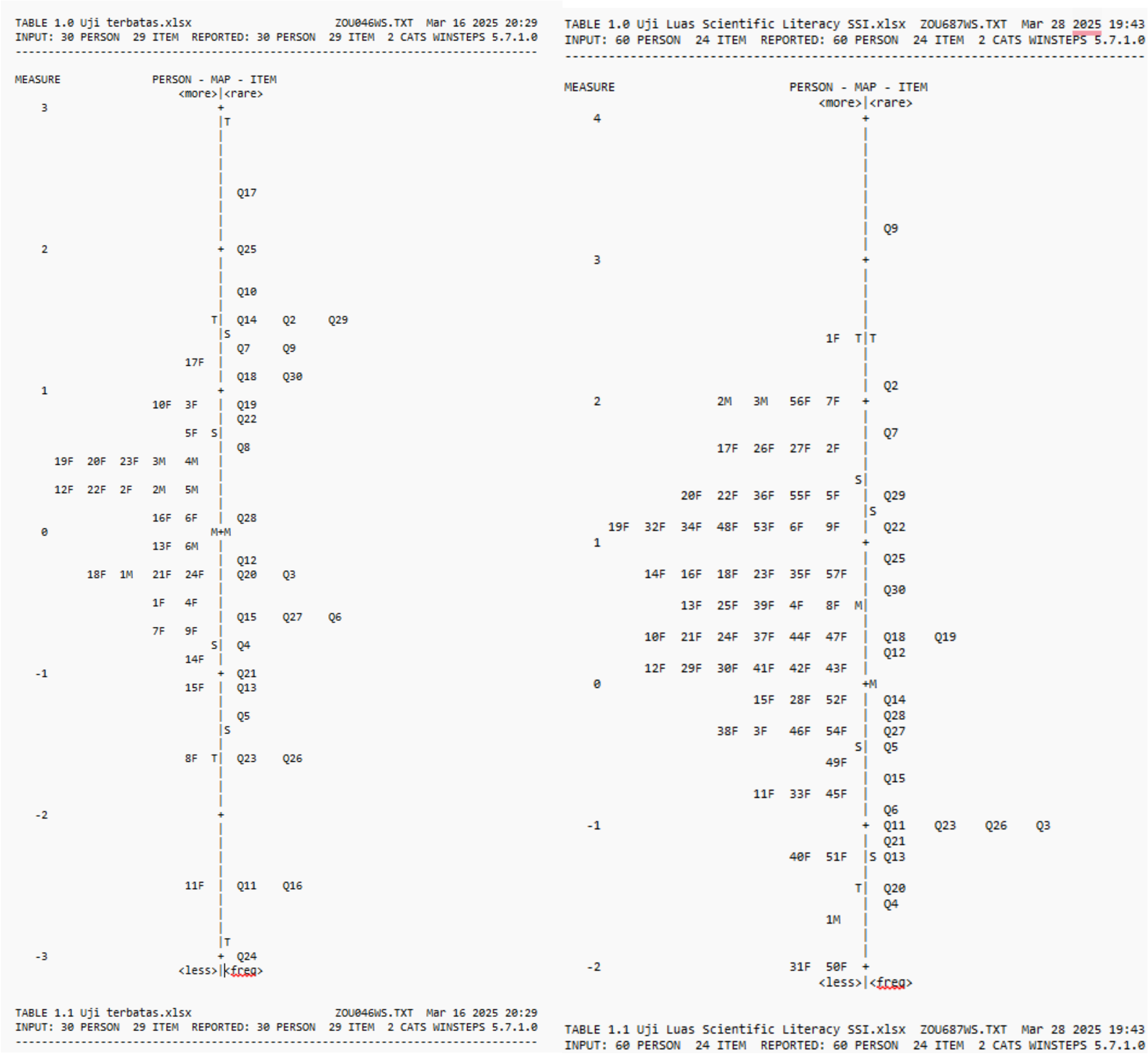

Figure 2. Difficulty Test Items Based on Variable (Wright) Map on Limited Sample (Left) and Wide Sample (Right)

Figure 2 shows that the test items have a "level" of difficulties. It shows that test items on level 4 are the hardest, which means the measure is above 1. On level 3 are the moderate test items, in which the measure is from above -1 to below +1. The level 2 test items are considered easy, which a range of measures from above -3 to below -1. Lastly, the easiest test items have a range of measures from below -3. The results of the difficulty test items from the variable (wright) map are shown in Table 9.

Table 9. The Difficulty Test items by Variable (Wright) Map on Limited Sample

Pilot Test	Level	Test item(s)	Frequency	Percentage (%)
Limited sample	Level 4	Q17, Q25, Q10, Q29, Q2, Q14, Q9, Q7, Q30, Q18	10	41.67
	Level 3	Q19, Q22, Q8, Q28, Q12, Q3, Q20, Q6, Q27, Q15, Q4	11	45.83
	Level 2	Q16, Q11	2	8.33
	Level 1	Q24	1	4.17

Pilot Test	Level	Test item(s)	Frequency	Percentage (%)
			Total	100
Wide sample	Level 4	Q9, Q2, Q7, Q29, Q22	5	20.83
	Level 3	Q25, Q30, Q18, Q19, Q12, Q14, Q28, Q27, Q5, Q15, Q6	11	45.83
	Level 2	Q11, Q23, Q26, Q3, Q21, Q13, Q20, Q4	8	33.33
	Level 1	-	0	0
			Total	100

The results show it can be observed that most of the test items are included on level 3 or moderate. While the easiest test items are the fewest with only one test items. Then, this research found out about the test items that were most difficult and the easiest to order. The difficulty of the test items has been descending while the research conducted the infit test items. It can be observed the difficulty of the test items on the right side. It can be concluded that the most difficult test items would be on the above and the easiest test items would be on the bottom. In the test items' difficulty analysis, four levels indicate the difficulty. It indicates that most of the test items in levels 3 and 4 that are considered difficult test items category have more numbers than test items in levels 1 and 2 that are considered easy test items.

Other analysis to find the quality of the test items, it required to analyze the discrimination of the test items. The discriminatory power for each test item for a limited and wide sample. To find the discriminatory power value, the number of the PT Measure Corr. is observed, and then the value is categorized into discriminatory power. The category of the discriminatory power in each test item is shown in Table 10.

Table 10. The Discriminatory Power in Each Test Item

Test item	PT Measure Corr. Value in Limited Sample	Category	PT Measure Corr. Value in Wide Sample	Category
Q2	0.18	Cannot discriminate	0.09	Cannot discriminate
Q3	0.15	Cannot discriminate	0.22	Fair
Q4	0.42	Very good	0.25	Fair
Q5	0.25	Fair	0.18	Cannot discriminate
Q6	0.12	Cannot discriminate	0.56	Very good
Q7	0.53	Very good	0.12	Cannot discriminate
Q8	0.33	Good	Invalid	Invalid
Q9	0.14	Cannot discriminate	0.07	Cannot discriminate
Q10	0.11	Cannot discriminate	Invalid	Invalid
Q11	0.50	Very good	0.31	Good
Q12	0.50	Very good	0.32	Good

Test item	PT Measure Corr. Value in Limited Sample	Category	PT Measure Corr. Value in Wide Sample	Category
Q13	0.08	Cannot discriminate	0.47	Very good
Q14	0.20	Fair	0.52	Very good
Q15	0.27	Fair	0.55	Very good
Q16	0.36	Good	Invalid	Invalid
Q17	0.28	Good	Invalid	Invalid
Q18	0.09	Cannot discriminate	0.44	Very good
Q19	0.18	Cannot discriminate	0.43	Very good
Q20	0.53	Very good	0.60	Very good
Q21	0.57	Very good	0.63	Very good
Q22	0.45	Very good	0.30	Good
Q23	0.27	Fair	0.59	Very good
Q24	0.34	Good	Invalid	Invalid
Q25	0.16	Cannot discriminate	0.51	Very good
Q26	0.20	Fair	0.39	Good
Q27	0.53	Very good	0.60	Very good
Q28	0.28	Fair	0.55	Very good
Q29	0.27	Fair	0.24	Fair
Q30	0.12	Cannot discriminate	0.18	Cannot discriminate

The test items analysis can be worked out similarly to find out the validity of the test items. Therefore, the test items that are not valid are analyzed. For example, the first pilot test had 29 test items and the second pilot test had 24 test items that were still included in the analysis before several test items decided to be valid or invalid. However, the test items that were invalid in the first pilot test were not included in the second pilot test. After the discrimination of the test items between the pilot test to limited and wide sample conducted, it needs to reveal the recapitulation to recap the results among the category of the test items discrimination analysis. The results of the recapitulation of the discrimination analysis of the test items are shown in Table 11.

Table 11. The Recapitulation of The Discriminatory Analysis

	Cannot Discriminate	Fair	Good	Very good	The sum of Test Items
Limited Sample	10	7	4	8	29
	34.48%	24.14%	13.79%	27.59%	
Wide Sample	5	3	4	12	24
	20.83%	12.50%	16.67%	50.00%	

The result of the recapitulation reveals the differences in the categories between the limited sample and the wide sample. It shows that the highest category in the pilot test of a limited sample cannot discriminate, while the lowest category is good. This result was contradictive to the result of the second pilot test, which revealed the highest category is very good, while the lowest category is fair. The result of discrimination shows that the recapitulation in the pilot test. The highest score is “cannot discriminate”. However, the test items are still acceptable because the values are not negative Ocy and Colleagues (2023) recommend conducting the inspection, whether it is to remove the test items or revise the options to get better value.

From the result, the pattern of the test items can be explained to differentiate between the test items that are considered difficult and easy. The test item that is considered difficult can be determined from the type of the test items, which all of them are complex multiple-choice. While the test item that is considered easy can be determined all of them are simple multiple-choice. The complex multiple-choice type is the type of test item that requires students to answer the question with more than one answer. If they only answer not completely all of the answers, then it counts as incorrect answers. It predicts that students are not familiar with these test items because usually answer for one answer in questions (Saputra et al., 2024). This type of complex multiple-choice is having two questions that correlate in one number. If the respondent has answered correctly in one of them, then it is considered incorrect in that number. It found that most of the students even prospective science teachers have difficulty in facing complex multiple-choice test items. This argument is supported by previous research that found the analysis of the scientific literacy achievement assessed to senior high school students that found the achievement of students in complex multiple-choice has a lower average than simple multiple-choice (Huryah et al., 2017).

Other previous research has been conducted on the development of the test items. Helendra and Sari (2021) conducted the development of the scientific literacy test items with similar topics regarding health and disease topics that consist of the excretory system and respiratory system, but this development of test items still follows the framework PISA 2015. However, the analysis of the developed test items is similar to this research. There is content validity which shows the average is 87.23%, and the construct validity of the development shows various results with the range 0.27 and 0.71 with a reliability is 0.93 included is high. In the difficulty level, that research is divided into three categories. Most of the test items show the medium category with 74% followed by the easy category with 14% and hard with 12%. The discriminatory index of that research is divided into four categories, which shows the highest category is good with 52%, followed by poor category with 24%, the third highest category is enough with 10%, and lastly very good category is the least with 14%.

Other research about the developing test items related to health and disease has similar steps of the development with this research was conducted by Wahyunisah and Susilawati (2023), that research developed about the scientific literacy test items in respiratory system topics. The content validity conducted by three experts shows the results are valid with a score of 82.5% and the construct validity that assesses 20 students for 30 test items shows there are 25 test items was valid. The reliability test results of the scientific literacy test instrument, as assessed by students, yielded a reliability coefficient of 0.67, which falls under the reliable category with a high classification. The difficulty level analysis of the test items indicated that 2 questions were categorized as easy, 20 questions as moderate, and 8 questions as difficult. Additionally, the item discrimination analysis revealed that 4 questions were classified as good, 11 questions as fair, and 15 questions as poor.

Azizah and Budijastuti (2021) has conducted the developing instrument abou the scientific literacy in the human circulatory system. It shows the similar steps of the development with another research. First, the researchers construct 15 test items about scientific literacy and ready to followe several steps. In the content validity, the rersearchers assess the aspect of the material, construct, and language to two experts which all of them is very valid. While in the construct of the validity, the researchers assess 20 high school students. The results of the validit shows 13 test items were valid and 2 test items were invalid, while the reliability analysis result is 0.722 which include on reliable. Lastly, the analysis that conducted by the researchers is about the difficulty of the test items which shows the hard test items are 2, medium test items are 11, and easy test items are 2.

Other researchers who would conduct similar research regarding the development of the scientific literacy test items need to have a foundation of the test items. To develop the test items, this research used the PISA 2025 framework with three competencies as a foundation, with some analysis of the students' textbooks to find out the weaknesses in those competencies. Therefore, the developed test items are the result of addressing the weaknesses.

The Profiling of Scientific Literacy on Students Based on Developed Test Items

The profiling of the test items was assessed based on similar characteristics of the participants. There were 194 PSTs collected from both the pilot test and profiling test assessed by 23 developed scientific literacy test items. This chapter reveals the analysis of the achievement of PST on all of the test items, themes of the test items, scientific literacy competencies and knowledge, and the SSI elements. The results show the graph among them by showing percentages of the achievement in each category. Also, the descriptive statistics that consist of the minimum, maximum, mean, and standard deviation regarding each result were revealed.

The first analysis is about the results that 194 PST achieved for 23 test items that have been developed to conduct the descriptive analysis. It shows the average of those participants and each participant toward the test items. Also, the result revealed about the most and the least among the PST. The descriptive statistics table provides an overview of a dataset which consists of 104 PST. The data exhibits a wide range of values, with a minimum of 4.35 and a maximum of 82.61, resulting in a range of 78.26. The mean (average) value is 62.58, indicating the central tendency of the dataset. However, the standard deviation of 15.33 suggests that the data points are moderately dispersed around the mean, meaning there is a considerable spread in the values. This is further confirmed by the variance of 235.053, which quantifies the overall dispersion of the dataset. The high variance and standard deviation indicate that while most values cluster around the mean, there are significant deviations, suggesting a diverse distribution of data points. The analysis of profiling to PSTs in each aspect, such as the competency, knowledge, SSI elements, and themes, was shown about the minimum, maximum, and mean of the achievement of the PSTs that shown in Table 12.

Table 12. The Achievement of PSTs in each Aspect of the Test Items

Aspect		Minimum	Maximum	Mean
Competency	Explain phenomena scientifically	14.43	85.05	58.66
	Construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically	9.79	83.51	63.77
	Research, evaluate and use scientific information for decision-making and action	20.62	80.93	57.82
Knowledge	Content	14.43	85.05	58.66
	Procedural	9.79	83.51	63.25
	Epistemic	20.62	80.93	58.42
SSI elements	Employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation	25.77	85.05	61.63
	Integrate implicit and/or explicit ethical components that require some degree of moral reasoning	14.43	82.99	57.39

	Aspect	Minimum	Maximum	Mean
	Emphasize the formation of virtue and character as long-range pedagogical goals	9.79	83.51	65.76
	Utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics	20.62	77.32	52.66
Theme	Corona virus	14.43	80.93	64.23
	Stunting	9.79	85.05	45.36
	Smoking	55.67	78.35	67.65
	Diabetes	33.51	83.51	65.57
	Acute respiratory system	25.77	78.87	55.77

The results reveal that competency 2 is “construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically” as the highest achievement with 63.77% of the PSTs answering correctly, while the second highest achievement by PST is achieved by competency 1 is “explain phenomena scientifically” with 58.66% answered by participants correctly. Lastly, the smallest competency achieved by PST is competency 3 “research, evaluate and use scientific information for decision-making and action”, achieved by about 57.82% of the participants answering correctly. This result is similar to the knowledge of scientific literacy. The result of the scientific literacy knowledge is almost similar to the competency. It found that the highest achievement is procedural knowledge, with 63.25% of the PST answering correctly. The second highest is the content knowledge is about 58.66% of the PST answers correctly. Lastly, the lowest achievement is epistemic knowledge, achieved by 58.42% of the PST.

The result reveals about the SSI elements that the highest achieved by the PST is SSI element is “emphasize the formation of virtue and character as long-range pedagogical goals,” with 65.76% of the PSTs answering correctly. The second-highest of the SSI elements achieved by PSTs is SSI element “employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation,” with 61.63%, and the third-highest is SSI element “integrate implicit and/or explicit ethical components that require some degree of moral reasoning” with 57.39% of the PSTs answering correctly. Lastly, the lowest SSI element achieved by the PST is SSI element “utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics,” with 52.66% of the PSTs answering correctly about this. Lastly, the achievement of the PSTs in the themes shows that the theme “smoking” is the highest with 67.65%, the second highest is “diabetes” with 65.57%, third highest of the theme is “corona virus” achieved about 64.23%. Lastly, the least achieved theme is about “stunting” with 45.36% of PSTs achieving this theme.

The result of the PSTs in University X comparing the three levels of the PSTs. The comparison of the level PSTs was conducted by analysing the average of the achievement in 23 test items, aspect of competency, knowledge, SSI elements, and themes PSTs from the first year, second year, and third year in university X. the results of the achievement of the PSTs for the 23 test items are shown in Figure 3.

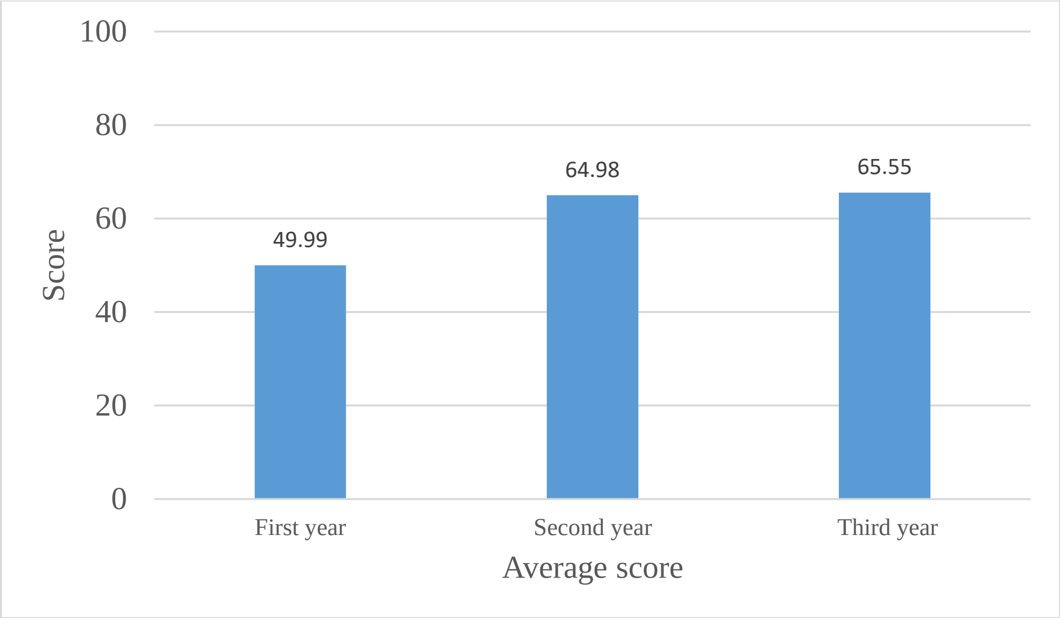

Figure 3. The Achievement of PSTs in each Year

Figure 3 shows that the average of the three levels of the PSTs, that shows third year PSTs is the highest and first year is the lowest. After that, the analysis of the independent sample t-test to measure whether the results are significantly different among three levels was conducted in Table 13.

Table 13. The Comparison Analysis between each Group

Comparison between the first year and the second year			
Group	**N**	**Sig. (2-tailed)**	**Decision**
First year	30	0.001	Significant different
Second year	19		
Comparison between the first year and the third year			
Group	**N**	**Sig. (2-tailed)**	**Decision**
First year	30	0.002	Significant different
Third year	13		
Comparison between the second year and the third year			
Group	**N**	**Sig. (2-tailed)**	**Decision**
Second year	19	0.915	Not significant different
Third year	13		

This analysis shows the results that show the achievement of the PSTs compared by each year. The table shows the result of the first-year PSTs mean score is the lowest among the other groups. The second-highest mean score was achieved by PSTS in the second year. The results between groups of PSTs in the first year and other years are 0.002, that less than 0.05, meaning that they are significantly different. While the comparison between the second year and third year shows 0.915, that more than 0.05, meaning that the results are not significantly different. Other analysis. Further analysis is about the achievement of the scientific literacy competency in each year, as shown in Figure 4.

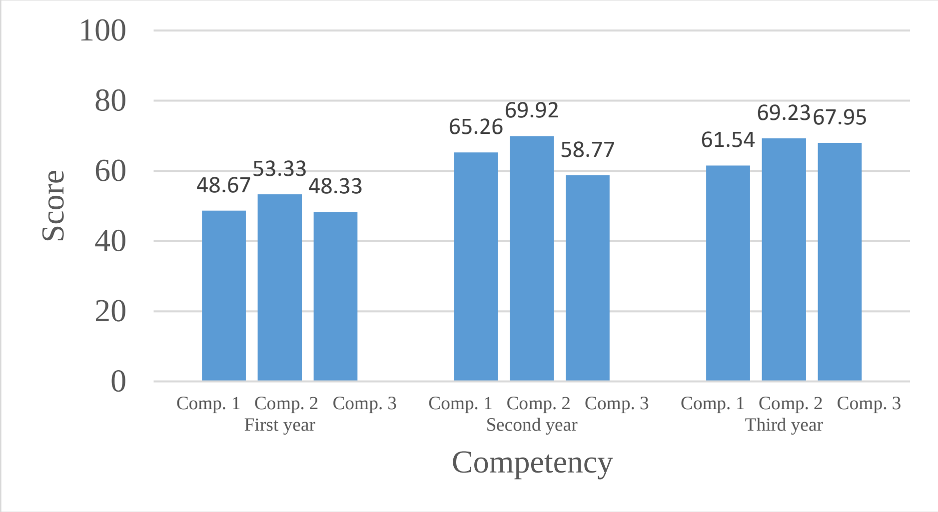

Figure 4. The Achievement of PSTs based on Competency of each Group

Figure 4 shows the third group about the year of the PSTs, in each year, there are three bar charts. The first bar is "comp. 1" is about the competency of "explain phenomena scientifically", the second bar is "comp.2" is about "construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically", and the third bar is "comp. 3" is about "research, evaluate and use scientific information for decision-making and action". PSTs in their first year have the highest achievement in the competency of "construct and evaluate designs for scientific enquiry, and interpret scientific data and evidence critically," and the lowest achievement in the competency of "research, evaluate and use scientific information for decision-making and action," similar as PSTs on second year. PSTs in the third year also have the same highest achievement in the competency with others, but the lowest achievement is in the competency of "explain phenomena scientifically". The other analysis is about the knowledge that was achieved by each group of PSTs shown in Figure 5.

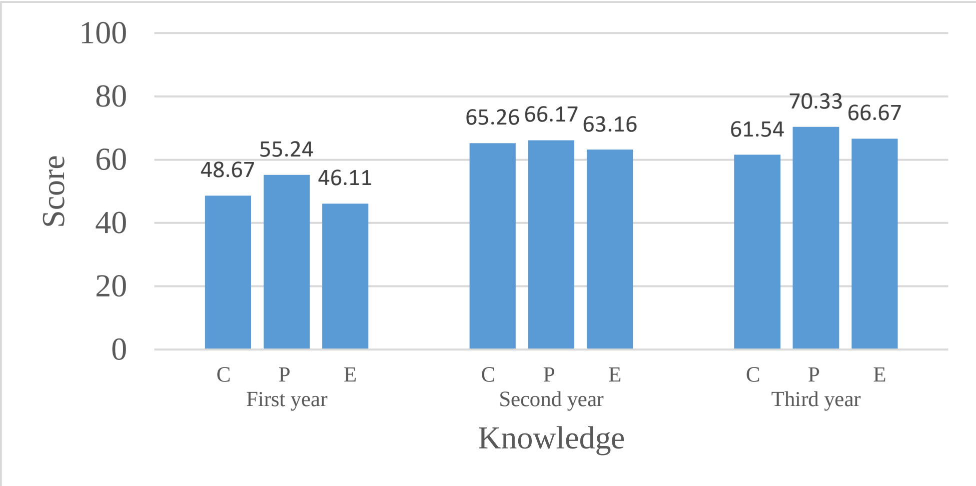

Figure 5. The Achievement of PSTs based on Knowledge in each Group

The achievement of PSTs based on the knowledge of scientific literacy consists of three bar charts in each group. The first bar is "C" about content knowledge, the second bar is "P" about procedural knowledge, and the third bar is "E" about epistemic knowledge. First-year PSTs highest achievement is procedural knowledge and the lowest achievement is epistemic same as PSTs in the second year. PSTs in the third year also have the highest achievement in procedural knowledge, but they have the lowest achievement in epistemic knowledge. Other analysis of the achievement by PSTs in each group based on the SSI elements is shown in Figure 6.

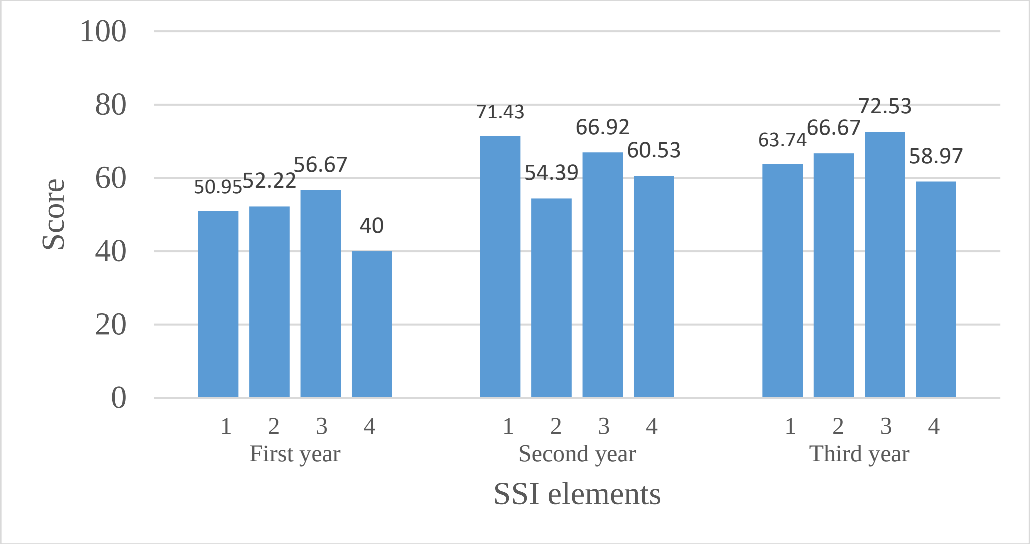

Figure 6. The Achievement of PSTs based on the SSI Elements

Figure 6 shows there are four bar charts in each group. The first bar is "employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation", the second bar is "integrate implicit and/or explicit ethical components that require some degree of moral reasoning", the third bar is "emphasize the formation of virtue and character as long-range pedagogical goals", and the fourth bar is "utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics".

The achievement by PST in the first year, the highest achievement based on SSI element is "emphasize the formation of virtue and character as long-range pedagogical goals" and the lowest is "utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics". in the second year PSTs, the highest achievement is "employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation" and the lowest is "integrate implicit and/or explicit ethical components that require some degree of moral reasoning". in the third year of PSTs, the highest SSI elements is "emphasize the formation of virtue and character as long-range pedagogical goals". Other analysis is about the achievement of PSTs based on the themes that are shown in Figure 7.

Figure 7 shows that five themes were assessed in PSTS. The first bar is about the theme of corona virus, the second bar is about stunting, the third bar is about smoking, the fourth bar is about diabetes, and the fifth bar is about the acute respiratory system. Each year, the achievements of the highest and lowest are different. The PSTs in the first year, the highest achievement is corona virus theme and the lowest achievement is acute respiratory infections. The result is different with the second year PSTs that the highest achievement is smoking theme and the lowest achievement is stunting theme same as achieved by third year PSTs. However, the PSTs in third year has highest achievement in the diabetes theme.

The result of the achievement in 194 PSTs conducted statisctics descriptive analysis. The average of the scores among PSTs is about 59.99, which is categorized as moderate achievement. The highest score achieved by PSTs is 86.96, and the lowest score achieved by PSTs is 4.35, and this difference makes a huge gap in the range score. The highest scientific literacy competency achieved by PST is "construct and evaluate designs for scientific enquiry and interpret scientific data and evidence critically" which is integrated into the inquiry skills. The comparison of the group on PSTs was conducted at the same university. First-year PSTs achieved the lowest average score compared to other years, followed by the achievement of the second-year PSTs as the second highest, and the highest achievement by the third year. There are significant differences in the score between the first year and the second year, and between the first year and the third year. However, the score between the second year and the third year shows no significant difference. This happens because third-year PSTs have more experience with the science learning with theory and practical activity for the whole semester than their junior.

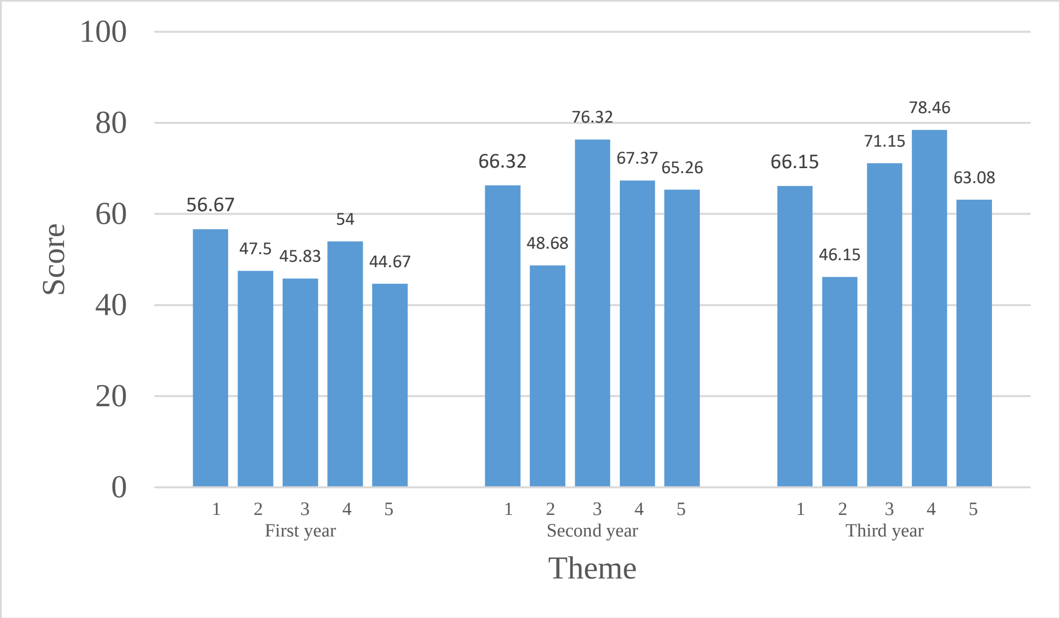

Figure 4. The Achievement of PSTs based on the Theme

The highest achievement in competency by PSTs is supported by the opportunity to conduct all of those activities to earn meaningful learning (Rustaman, 2005). Therefore, this competency is correlated with the procedural knowledge and epistemic knowledge, one of which has a higher average than content knowledge. The inquiry skills and the ability the interpret data, that consider also as science process skills, were gained from the experiences (Rustaman, 2007). Therefore, third-year PSTs could achieve higher than their lower year. This is proven by Mijaya and Colleagues (2019) show that the aspect before treatment is lower than after treatment. It reflects the competency in the inquiries and interpreting data gained from the experiences. This argument is supported by the research that found the results indicate that 24% of students possess scientific process skills at a moderate level, while 76% fall into the low category. When analyzing specific aspects of these skills, students were found to struggle particularly with formulating hypotheses, identifying variables, and designing experimental procedures. However, they demonstrated a fair level of competence in selecting appropriate tools and materials. The overall low proficiency in scientific process skills is attributed to a lack of adequate training and practice in daily learning activities (Mahmudah et al., 2019). Other caused that might influence are limitations in practical session time, prior training in skills, opportunities to test equipment, and the clarity of information received by students (Adiningsih et al., 2019). It strengthens the argument that the PST in this case has experiences about improving science process skills because they focused on the subject of science, rather than the students that have other subject besides science and have limitation of the time if they conduct the activity.

The second-highest competency achieved by PST is “explain phenomena scientifically”. This competency requires students to use their knowledge of the content. This competency indicates the ability of the students is aligned to the concept formation and understanding of the students (Ogundeji et al., 2020). This result is aligned to content knowledge because this competency is focused on the explaining scientific, technological, and environmental phenomena goes beyond simply recalling and applying theories, concepts, and factual knowledge. It also requires an understanding of how this knowledge is obtained and the degree of confidence we can place in scientific claims (OECD, 2023). This result also considered moderate results because more than 50% of the participants were able to answer the test items correctly.

The lowest competency achieved by PST is “research, evaluate and use scientific information for decision-making and action”. This competency is the newest in the PISA scientific literacy competency. The research by Kinskey and Zeidler (2021) about the decision-making skills possessed by students needs to be able to have the knowledge or completely understand the concept. Therefore, this competency is the lowest among PST because it needs a long process from the research to make a decision and all of those processes must be correct and align to the competency. This shows the possible correlation between the nature of science that relates to the term of “research” in the competency to “decision-making” because they justify their decisions based on the understanding of the empirical, tentative, and subjective and emphasize the aspect of nature of science in the science topics before making a decision (Khishfe, 2012). Decision-making skills also have positive correlation to problem-solving skills. Also, there was a moderate positive correlation between students' decision-making skills and their confidence in problem-solving abilities. Additionally, a low but significant positive correlation was found between the self-control and avoidance sub-dimensions (Yurtseven et al., 2014).

The results of this research align with previous research that assessed the scientific literacy of prospective elementary teachers and found that the highest competency is “construct and evaluate designs for scientific enquiry” with 50% and the second highest competency is “explain phenomena scientifically” with 43%. Lastly, the lowest competency achieved by the competency of “interpret scientific data and evidence critically” with a 27% achievement (Fauziah et al., 2024). Other research that assessed the same level of prospective elementary teachers, found that the highest competency is “explain phenomena scientifically” with that has average score is 55, while the competency “construct and evaluate designs for scientific enquiry” has an average score of 49, and the lowest competency that achieves is “interpret scientific data and evidence critically” has average score 49 (Mulyani & Zulkarnaen, 2024). While assessing prospective chemistry teachers by the PISA 2015 scientific literacy framework, it found that the highest competency achieved by them is “explain phenomena scientifically” with 52.5% achieved. While the second highest competency is “construct and evaluate designs for scientific enquiry” which was achieved by 52%. The smallest competency achieved by them is “interpret scientific data and evidence critically” which is achieved by 40% of the population (Suryanti et al., 2020).

Other cases that influence the knowledge that correlates with the competency argue by Kranz and Colleagues (2023) state that challenges are: (1) Learners often struggle with understanding the scientific concepts that underlie different phases of inquiry and experimental activities. These challenges include difficulties such as not knowing what a hypothesis is, failing to control variables, conducting experiments without a control condition, and confusing observation with data interpretation. Additionally, learners may struggle to make exclusion inferences when analyzing results. Challenges in the epistemic domain stem from difficulties in understanding the purpose of experimentation and constructing explanations, especially given that scientific knowledge evolves with new evidence or reinterpretations of existing data. Some of these challenges indicate that learners struggle with conducting scientific investigations due to a lack of procedural skills. Engaging with real-life experimental objects, such as equipment, chemicals, or living organisms in biology can help students develop practical experience. However, difficulties often arise during hands-on activities and experiment execution. These challenges include handling materials and substances carelessly, repeatedly conducting the same trials without the intent of obtaining repeated measurements, and failing to incorporate all necessary test trials in their experiments. Nonetheless, evidence of such procedural challenges remains limited, as these issues were relatively infrequent in the current review.

PSTs in the third year achieved the highest knowledge of procedural and epistemic knowledge than the first and second year of PSTs. All of the groups of the PSTs achieved the procedural knowledge as the highest achievement. Procedural knowledge, understanding the standard procedures and methods that scientists use to gather reliable and valid data is essential. This knowledge is necessary not only for conducting scientific investigations but also for critically evaluating the evidence used to support conclusions drawn from that data. While epistemic knowledge is an understanding of the key constructs and defining characteristics fundamental to how knowledge is built in science, as well as their role in validating scientific knowledge. Epistemic knowledge thus offers the reasoning behind the methods and practices scientists use, outlines the structures and core features that direct scientific investigation, and serves as the foundation for trusting the claims science makes about the natural world. (OECD, 2023). Also, PSTs in third year has more experiences about the epistemic knowledge because they have experiences about nature of science or science learning.

The SSI element of “emphasize the formation of virtue and character as long-range pedagogical goals” is the highest achievement by 194 PSTs. However, first-year PSTs scored 56.67, indicating a foundational awareness of the importance of character and virtue in science education. This score increases significantly in the second year to 66.92, suggesting that as PSTs progress through their training, they begin to more fully understand and integrate these long-range educational goals into their planning and instruction. By the third year, PSTs achieved the highest score of 72.53, reflecting a mature pedagogical perspective that emphasizes not just content knowledge, but also the cultivation of ethical and virtuous learners. This upward trend suggests that teacher education programs are effective in gradually instilling the importance of character education. Moreover, it indicates that extended exposure to teaching practice and reflection likely enhances PSTs' ability to see science education as a vehicle for shaping responsible, morally grounded citizens. These findings are supported by the argumentation by Carr (2017) state at the initial stage of development, young individuals are not yet capable of exercising the practically wise deliberation characteristic of a more morally or virtuously mature agent; rather, they must acquire such discernment through a process of education and experiential growth. In addition, there is impact of the curriculum intervention to understanding and practice of virtue, that they acknowledgeg about the human beings about what is the things that is right or is necessary for character education (Pike et al., 2020)

In the SSI elements of “employ the use of scientific topics with social ramifications that require students to engage in dialogue, discussion, debate, and argumentation” is the second-highest achievement by 194 PSTs. It supports that by previous research, especially about their argumentation regarding scientific topics that most students have low quality of argumentation skills because most of them are still in the phase of claim, data, and warrant without fulfilling the phase of backing and rebuttal (Taufik et al., 2023). It was also found in a higher performing group of argumentation, they exchanged their knowledge, beliefs, and values to develop a mutual understanding and reach a consensus. In contrast, while low-performing groups attempted to align through argumentative discourse, their lack of critical evaluation, reasoning, or constructive debate led to quick, superficial agreements, ultimately hindering their creative outcomes (Hernandez Sibo et al., 2023). The scientifically literate person also needs to have their engagement and criticism by the argumentation about news science reports from various sources. Therefore, the quality of the argumentation align to the quality of the critical thinking of the person (Lin, 2014). First-year PSTs achieved the lowest rather than group PSTs. Their prior education may have prioritized memorization over critical discourse, leaving them unprepared for the open-ended nature of debates on topics like public health policies, government programmes to reduce stunting, and others. Additionally, first-year students are often still developing the interpersonal and rhetorical skills needed for productive dialogue, including active listening, respectful disagreement, and persuasive communication.

The third highest achievement in SSI elements section by PSTs is “integrate implicit and/or explicit ethical components that require some degree of moral reasoning”. The results of the comparison, the third year PSTs achieved as the highest rather than in other years, because they are the most mature and oldest. the moral reasoning helps people to change society when it is needed. To argue the change, individuals need to argue about the things that are unacceptable or unfair. Other, to create the change in society, individuals need to critically examine the limitations of prevailing attitudes, including stereotypes and cognitive biases (Killen & Dahl, 2021). Third-year PSTs possess the highest because they have more extensive academic and practical experiences in higher education. While the first-year PSTs are the lowest, they are still in early stages of training their moral reasoning skills, especially about society. Additionally, first-year students may lack exposure to diverse perspectives, limiting their ability to empathize or consider ethical implications beyond personal biases.

The result of the SSI elements shows that the “utilize personally relevant, controversial, and ill-structured problems that require scientific, evidence-based reasoning to inform decisions about such topics” possess the lowest achievement. It is because PST face some difficulties such as facing controversial issues and they need to decide about their positions. Also, besides implementing the knowledge that the participants have, the SSI emerged the participants to consider when they make decisions based on the case by considering the ethics/morality, sociology/culture, economy, and policy (Cebesoy & Rundgren, 2023). The decision-making skills are influenced by many variables. The research found by (Álvarez-Justel & Ruiz-Bueno, 2021), that decision-making skills are influenced proportionally to several dimensions such as emotional, rational, intuitive, cognitive, dependent, social, and academic (Álvarez-Justel & Ruiz-Bueno, 2021). The decision-making skills were a complex process. The current approach, which combines conceptual understanding with inquiry-based analysis of evidence, value clarification, and argumentation, offers a foundational framework for enhancing efforts to cultivate students' scientific literacy and their ability to make rational decisions in a socio-scientific context (Lee, 2007). The PSTs on first year achieved the lowest in this SSI elements, because they may not yet to see the connection between the controversion things. Also, their prior education has often emphasized structured problems with clear solutions, leaving them unprepared for the ambiguity of real-world dilemmas, that cause lack of the critical thinking experiences about the complex issues. The topics of stunting expressed as the lowest achievement among of the PSTs in general.

CONCLUSION

The 23 constructed valid test items, as the final instrument of this research, from 30 test items as the initial, have been subjected to several tests to ensure the quality of the instrument with high reliability, the distribution of the difficulty, and the discrimination of the test items by two pilot tests with different number of samples. Additionally, the content also ensures the quality of the scientific literacy test items by expertise in the assessment and content field, resulting in good validity of the test items, and follows the readability test on the PSTs to refine the language accuracy and clarity. The final test items conducted during the large-scale implementation with 194 PSTs provided insights into their scientific literacy levels, showing moderate results, with competency 2, procedural knowledge, SSI element 3, and smoking theme being the highest achievement. The lowest achievement by PSTs is about competency 3, epistemic knowledge, SSI element 4, and the stunting theme. The comparison among the first, second, and third years of PSTs was conducted, which shows the first and second years, and the first and third years are significantly different, but the comparison of the second and third year is not significant different.

The findings have significant implications for science education, particularly in improving how socioscientific issues are integrated into scientific literacy instruction. The lack of emphasis on decision-making competencies in textbooks suggests that students may struggle to apply scientific knowledge in real-world contexts. This indicates a need for instructional approaches that go beyond factual learning and focus on critical thinking, argumentation, and ethical reasoning. The study also highlights the importance of well-structured assessments that measure not only content knowledge but also higher-order cognitive skills.

RECOMMENDATION

To address these challenges, science curricula should incorporate a more balanced representation of all scientific literacy competencies, ensuring that decision-making and inquiry-based learning are given as much emphasis as content knowledge. Textbooks and instructional materials should be revised to integrate SSI elements that challenge students to critically evaluate scientific information. In addition, science assessments should move beyond knowledge recall to include reasoning, argumentation, and decision-making skills, with the use of open-ended and case-based questions providing a more comprehensive evaluation of students' scientific literacy. Teacher preparation programs should also focus on equipping educators with effective strategies for teaching socioscientific issues, including techniques for facilitating classroom discussions, integrating ethical considerations, and promoting evidence-based reasoning. Future research should explore the long-term impact of SSI-based instruction on students’ reasoning and decision-making skills, particularly through longitudinal studies assessing whether this approach leads to improved scientific engagement in higher education and professional settings. Also, future research needs to engage the teachers’ efficacy after being assessed by scientific literacy test items. Additionally in the development stage, the future research should collect the validity evidence such as the convergent validity with established scientific reasonigs test and response process analysis.

ACKNOWLEDGMENT

Thank you for prospective science teachers from Universitas Pendidikan Indonesia, Universitas Negeri Malang, and Universitas Negeri Yogyakarta for their effort and willingness for this research.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Winata Tegar Saputra	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Lilit Rusyati	✓					✓		✓	✓	✓		✓		
Nuryani Y. Rustaman	✓	✓		✓		✓					✓	✓		
Indra Fardhani			✓		✓		✓				✓			
I. Fikri Natadiwidjaya		✓			✓		✓		✓	✓				

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

Adawiyah, R., & Wisudawati, A. W. (2017). Pengembangan Instrumen Tes Berbasis Literasi Sains : Menilai Pemahaman Fenomena Ilmiah Mengenai Energi. *Indonesian Journal of Curriculum*, *5*(2), 112–121.

Adiningsih, M. D., Karyasa, I. W., & Muderawan, I. W. (2019). Profil Keterampilan Proses Sains Siswa Dalam Praktikum Titrasi Asam Basa. *Jurnal Pendidikan Kimia Indonesia*,

3(2), 94. https://doi.org/10.23887/jpk.v3i2.21241
Aini, S. N., & Pramasdyahsari, A. S. (2023). Pendampingan Simulasi Asesmen Kompetensi Minimum berbasis Literasi Digital untuk Meningkatkan Kemampuan Literasi Numerasi dan Adaptasi Teknologi Siswa Sekolah Dasar. *Jurnal Pengabdian Masyarakat Bangsa*, *1*(6), 573–583. https://doi.org/10.59837/jpmba.v1i6.210
Álvarez-Justel, J., & Ruiz-Bueno, A. (2021). Decision-making Profiles Secondary-school Students. *RELIEVE*, *27*(1).
Astuti, W. P., Prasetyo, A. P. B., & Rahayu, E. S. (2014). Pengembangan Instrumen Asesmen Autentik Berbasis Literasi Sains Pada Materi Sistem Ekskresi. *Lembaran Ilmu Kependidikan*, *43*(2), 94–102.
Azizah, Z. N., & Budijastuti, W. (2021). Pengembangan Instrumen Penilaian untuk Mengukur Keterampilan Literasi Sains pada Submateri Sistem Peredaran Darah Manusia. *Berkala Ilmiah Pendidikan Biologi (BioEdu)*, *11*(1), 89–97. https://doi.org/10.26740/bioedu.v11n1.p89-97
Branch, R. M. (2009). Instructional Design: The ADDIE Approach. In *Springer Science BusinessMedia, LLC* (Vol. 7, Issue 2). Springer New York Dordrecht Heidelberg London. https://doi.org/10.1007/978-0-387-09506-6
Carr, D. (2017). Virtue and Character in Higher Education. *British Journal of Educational Studies*, *65*(1), 109–124. https://doi.org/10.1080/00071005.2016.1224806
Cebesoy, U. B., & Rundgren, S.-N. C. (2023). Embracing Socioscientific issues-based Teaching and Decision-making in Teacher Professional Development. *Educational Review*, *75*(3), 507–534. https://doi.org/10.1080/00131911.2021.1931037
Chasanah, N., Widodo, W., & Suprapto, N. (2022). Pengembangan Instrumen Asesmen Literasi Sains Untuk Mendeskripsikan Profil Peserta Didik. *PENDIPA Journal of Science Education*, *6*(2), 474–483. https://doi.org/10.33369/pendipa.6.2.474-483
Creswell, J., & Creswell, D. (2023). Research Design, Qualitative, Quantitative and Mixed Methods Approaches. In *SAGE Publications,Inc.: Vol. Sixth Edit* (Issue 1). https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf
Dewi, N. L. W. A., & Putra, I. K. D. A. S. (2022). Analisis Aspek Literasi Sains Pada Buku Tematik Pembelajaran Ipa Kelas Vi Di Sd Negeri 2 Cempaga Tahun 2022. *ORBITA: Jurnal Pendidikan Dan Ilmu Fisika*, *8*(2), 339. https://doi.org/10.31764/orbita.v8i2.11478
Faisal, & Martin, S. N. (2022). Exploring Indonesian Biology Teachers' Perceptions and Attitudes Towards Socio-Scientific Issues-Based Instruction. *Asia-Pacific Science Education*, *8*(1), 256–291. https://doi.org/10.1163/23641177-bja10042
Fauziah, N., Ningsyih, S., & Khusaini, F. (2024). Profil Kemampuan Literasi Sains Mahasiswa Program Studi Pendidkan Guru Sekolah Dasar Pada Mata Kuliah Pendidikan Lingkungan Hidup : Sebuah Studi Pendahuluan. *Journal of Classroom Action Research*, *6*(1). https://jppipa.unram.ac.id/index.php/jcar/article/view/6971
Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2017). How to Design and Evaluate Research in Education. In *McGraw-Hill* (Vol. 01).
Genisa, M. U., Subali, B., Djukri, Agussalim, A., & Habibi, H. (2020). Socio-Scientific Issues Implementation as Science Learning Material. *International Journal of Evaluation and Research in Education*, *9*(2), 311–317. https://doi.org/10.11591/ijere.v9i2.20530
Hasanah, K., Siregar, L. F., & Sumanik, N. B. (2024). Rancang Bangun Instrumen Soal Literasi Sains Berbasis Quizizz Pada Pokok Bahasan Larutan Elektrolit Dan Non Elektrolit. *LENSA (Lentera Sains): Jurnal Pendidikan IPA*, *14*(1), 58–68. https://doi.org/10.24929/lensa.v14i1.497
Helendra, H., & Sari, D. R. (2021). Pengembangan Instrumen Asesmen Berbasis Literasi Sains Tentang Materi Sistem Ekskresi dan Sistem Pernapasan. *Jurnal Ilmiah Pendidikan Profesi Guru*, *4*(1), 17–25. https://doi.org/10.23887/jippg.v4i1.29860
Hernandez Sibo, I. P., Gomez Celis, D. A., Liou, S., Koh, B., & Leung, A. K. Y. (2023). The

Role of Argumentation on High- and Low-Creative Performing Groups: A Structuration Analysis of Undergraduate Students' Group Discussion. *Thinking Skills and Creativity*, *47*, 1–13. https://doi.org/10.1016/j.tsc.2022.101217
Huryah, F., Sumarmin, R., & Effendi, J. (2017). Analisis Capaian Literasi Sains Biologi Siswa Sma Kelas X di kota Padang. *Jurnal Eksakta Pendidikan (Jep)*, *1*(2), 72. https://doi.org/10.24036/jep.v1i2.70
Khishfe, R. (2012). Nature of Science and Decision-Making. *International Journal of Science Education*, *34*(1), 67–100. https://doi.org/10.1080/09500693.2011.559490
Killen, M., & Dahl, A. (2021). Moral Reasoning Enables Developmental and Societal Change. *Perspectives on Psychological Science*, *16*(6), 1209–1225. https://doi.org/10.1177/1745691620964076
Kinskey, M., & Zeidler, D. (2021). Elementary Preservice Teachers' Challenges in Designing and Implementing Socioscientific Issues-Based Lessons. *Journal of Science Teacher Education*, *32*(3), 350–372. https://doi.org/10.1080/1046560X.2020.1826079
Kranz, J., Baur, A., & Möller, A. (2023). Learners' Challenges in Understanding and Performing Experiments: A Systematic Review of The Literature. *Studies in Science Education*, *59*(2), 321–367. https://doi.org/10.1080/03057267.2022.2138151
Kusuma, A. S., & Nurmawanti, I. (2023). Pengembangan Soal-Soal Literasi dan Numerasi Berbasis High Order Thinking Skills (HOTS) untuk Siswa Sekolah Dasar (SD). *Jurnal Ilmiah Profesi Pendidikan*, *8*(1), 516–523. https://doi.org/10.29303/jipp.v8i1.1313
Latif, A., Pahru, S., & Muzakkar, A. (2022). Studi Kritis Tentang Literasi Sains dan Problematikanya di Sekolah Dasar. *Jurnal Basicedu*, *6*(6), 9878–9886. https://doi.org/10.31004/basicedu.v6i6.4023
Lee, Y. C. (2007). Developing Decision-Making Skills for Socio-Scientific Issues. *Journal of Biological Education*, *41*(4), 170–177. https://doi.org/10.1080/00219266.2007.9656093
Lin, S. S. (2014). Science and Non-Science Undergraduate Students' Critical Thinking and Argumentation Performance in Reading A Science News Report. *International Journal of Science and Mathematics Education*, *12*(5), 1023–1046. https://doi.org/10.1007/s10763-013-9451-7
Mahmudah, I. R., Makiyah, Y. S., & Sulistyaningsih, D. (2019). Profil Keterampilan Proses Sains (KPS) Siswa SMA di Kota Bandung. *Jurnal Diffraction*, *1*(1), 39–43.
Martinah, A. A., Mubarok, V., Miarsyah, M., & Ristanto, R. H. (2022). Pengembangan Instrumen Tes Literasi Sains Berbasis Kontekstual pada Materi Pencemaran Lingkungan. *Bioedusiana: Jurnal Pendidikan Biologi*, *6*(2), 192–218. https://doi.org/10.37058/bioed.v6i2.3251
Maulida, F., & Sunarti, T. (2022). Pengembangan Instrumen Tes Literasi Sains Berbasis Kearifan Lokal Di Kabupaten Lamongan. *ORBITA: Jurnal Pendidikan Dan Ilmu Fisika*, *8*(1), 52. https://doi.org/10.31764/orbita.v8i1.8337
Mijaya, N. P. A. P., Sudiatmika, A. A. I. A. R., & Selamet, K. (2019). Profil Literasi Sains Siswa Smp Melalui Model Pembelajaran Levels of Inquiry. *Jurnal Pendidikan Dan Pembelajaran Sains Indonesia (JPPSI)*, *2*(2), 161. https://doi.org/10.23887/jppsi.v2i2.19385
Mulyani, H., & Zulkarnaen, R. H. (2024). Profil Kemampuan Literasi Sains Calon Guru SD: Analisis Pada Aspek Kompetensi. *Pendas: Jurnal Ilmiah Pendidikan Dasar*, *09*(02), 304–316.
Murti, W. W., & Sunarti, T. (2021). Pengembangan Instrumen Tes Literasi Sains Berbasis Kearifan Lokal Di Trenggalek. *ORBITA: Jurnal Kajian, Inovasi Dan Aplikasi Pendidikan Fisika*, *7*(1), 33. https://doi.org/10.31764/orbita.v7i1.4386
Nabil, N. R. A., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis Indeks Aiken Untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia. *Jurnal Penelitian Pendidikan*, *25*(2), 184–191.

Nida, S., Rahayu, S., & Eilks, I. (2020). A Survey of Indonesian Science Teachers' Experience and Perceptions toward Socio-Scientific Issues-Based Science Education. *Education Sciences*, *10*(2), 1–15. https://doi.org/10.3390/educsci10020039

Nurfaidah, S. S. (2017). Analisis Aspek Literasi Sains Pada Buku Teks Pelajaran Ipa Kelas V Sd. *Mimbar Sekolah Dasar*, *4*(1), 56–66. https://doi.org/10.53400/mimbar-sd.v4i1.5585

Ocy, D. R., Rahayu, W., & Makmuri, M. (2023). Rasch Model Analysis: Development of Hots-Based Mathematical Abstraction Ability Instrument According To Riau Islands Culture. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *12*(4), 3542. https://doi.org/10.24127/ajpm.v12i4.7613

OECD. (2007). PISA 2006, Science Competencies for Tomorrow's World Volume1: Analysis. In *PISA, OECD Publishing, Paris* (Vol. 1). http://www.oecd-ilibrary.org/education/pisa-2006_9789264040014-en

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education: Vol. I*. http://dx.doi.org/10.1787/9789264266490-en

OECD. (2023). *Pisa 2025 Science Framework*. *May 2023*, 1–93.

Ogundeji, O. M., Madu, B. C., & Onuya, C. C. (2020). Scientific Explanation of Phenomena and Concept Formation as Correlates of Students' Understanding of Physics Concepts. *Journal of Natural Sciences Research*, *10*(3), 10–19. https://doi.org/10.7176/jnsr/11-16-03

Pike, M. A., Hart, P., Paul, S. S., Lickona, T., & Clarke, P. (2020). Character Development Through The Curriculum: Teaching and Assessing The Understanding and Practice of Virtue. *Journal of Curriculum Studies.*, 20.

Putri, R. K. (2020). Pengembangan Instrumen Tes Literasi Sains Siswa Pada Topik Keanekaragaman Makhluk Hidup. *Diklabio: Jurnal Pendidikan Dan Pembelajaran Biologi*, *4*(1), 71–78. https://doi.org/10.33369/diklabio.4.1.71-78

Rohim, D. C., Rahmawati, S., & Ganestri, I. D. (2021). Konsep Asesmen Kompetensi Minimum untuk Meningkatkan Kemampuan Literasi Numerasi Siswa Sekolah Dasar. *Jurnal VARIDIKA*, *33*(1), 54–62. https://doi.org/10.23917/varidika.v33i1.14993

Rosidah, F. E., & Sunarti, T. (2017). Pengembangan Tes Literasi Sains pada Materi Kalor di SMA Negeri 5 Surabaya. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, *06*(03), 250–257.

Rustaman, N. Y. (2005). Development of Research in Inquiry Science Learning. *Seminar Nasional II*, 22–23.

Rustaman, N. Y. (2007). Keterampilan Proses Sains. In *Universitas Pendidikan Indonesia, Bandung*.

Saputra, W. T., Rustaman, N. Y., & Rusyati, L. (2024). Evaluating Validity and Reliability of Scientific Literacy Among Indonesian Secondary Students in PISA 2015 Context : Rasch Model Analysis. *Journal of Research and Educational Research Evaluation*, *13*(2), 169–179.

Sholikhah, F. F., & Purwani, A. T. (2023). Konsep Assesmen Kompetensi Minimum Dalam Meningkatkan Kemampuan Literasi Numerasi Peserta Didik. *Al-Ikmal: Jurnal …*, *2*(2), 27–33. https://e-journal.iaidalampung.ac.id/index.php/al-ikmal/article/view/101

Suryanti, S., Nufida, B. A., Hendrawani, H., & Rahayu, S. (2020). Identification of Science Literacy Competency of Chemistry Teacher Candidate. *Jurnal Zarah*, *8*(1), 50–55.

Taufik, A. N., Berlian, L., Iman, A., & Tarisa, R. (2023). Profil Kemampuan Argumentasi Lisan Berbasis Socioscientific Issues Mahasiswa Pendidikan IPA pada Mata Kuliah Mikrobiologi. *PENDIPA Journal of Science Education*, *6*(3), 832–838. https://doi.org/10.33369/pendipa.6.3.832-838

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

Wahyunisah, R., & Susilawati. (2023). Pengembangan Instrumen Tes Literasi Sains pada Materi Sistem Pernapasan Manusia Untuk Siswa SMP Kelas VIII. *Journal of Natural*

Science Learning, *2*(1), 43–51.

Yurtseven, R., Baysal, Ü. E. A., & Ocak, G. (2014). Analysis of the Relationship Between Decision Making Skills and Problem Solving Skills of Primary School Students. *International Online Journal of Education and Teaching (IOJET)*, *8*(3), 2148–2225.

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