26 (3), 2025, 1377-1405
Jurnal Pendidikan MIPA
e-ISSN: 2685-5488 | p-ISSN: 1411-2531
https://jpmipa.fkip.unila.ac.id/index.php/jpmipa

Development of a Science Literacy Test for Junior High School Students Based on the PISA 2025 Framework

Farah Amara Valio, Ledya Safira, Uunwanah Agustin Aulliyah, & Iwan Permana Suwarna*
Department of Physics Education, Syarif Hidayatullah State Islamic University, Indonesia

Abstract: This study aims to develop a valid, reliable, and practical science literacy test instrument based on the PISA 2025 framework as a tool for assessing the science literacy abilities of junior high school students in a contextual and in-depth manner. The issue of low science literacy in Indonesia, where 53.60% of students are in the very low category, highlights the urgency of providing an evaluation instrument that can comprehensively represent scientific thinking abilities. This achievement is closely related to the lack of student training using international assessment-based testing instruments such as PISA during the learning process. The research method used was Research and Development (R&D) with a 4-D development model (Define, Design, Development, Dissemination). The instrument was developed based on four dimensions of science literacy (competence, context, knowledge, and cognitive level) within the PISA 2025 framework and covers topics in the science subject, namely electricity, waves, and magnetism. Content validation was conducted by content, instrument, and language experts, while empirical testing was carried out to evaluate the quality of the test items. The research results showed that the instrument had high content validity (CVI = 0.93; CVR = 0.8–1). The average validity of each item reached 0.60. The instrument showed good consistency in terms of reliability, as indicated by the Omega McDonald coefficient of 0.79 for the combination of essay and complex multiple-choice questions, and the Cronbach's Alpha value of 0.68 for multiple-choice questions analysed separately. Most items were categorized as moderately difficult (72.73%) and had good discriminative power (63.63%). Additionally, the practicality value of 78.85% indicates that the instrument is easy to use in an educational context. Therefore, this instrument is suitable for use as a science literacy assessment tool aligned with the PISA 2025 framework and supports the development of higher-order thinking skills in national assessments.

Keywords: science literacy, instruments, assessment, PISA.

▪ INTRODUCTION

The scientific literacy skills of Indonesian students face serious challenges in the context of global competition. According to the 2022 PISA results, Indonesia ranked 64th out of 81 countries with a science literacy score of 383, a decrease of 13 points from 2018 (396) and far below the OECD average (485) (OECD, 2023a). This consistent decline from 403 (2015) to 396 (2018) to 383 (2022) indicates the urgency of fundamental reforms in science education and assessment. Science literacy is not merely an academic indicator but a critical foundation for developing scientific problem-solving skills and global competitiveness in the 21st century (Costa, Loureiro, & Ferreira, 2021; Elhai, 2023; Townley, 2018). The challenges of the 21st century require innovative science-based solutions, which demand the strengthening of higher-order thinking skills (HOTS) such as analysis, evaluation, and creativity in a scientific context (Nisak & Yulkifli, 2021; Zulfiani, Permana Suwarna, Muin, Mulyati, & El Islami, 2023).

The low science literacy scores of Indonesian students indicate deep-rooted problems related to the implementation of science education in schools. This situation

Iwan Permana Suwarna
*Email:
iwan.permana.suwarna@uinjkt.ac.id

DOI: http://dx.doi.org/10.23960/jpmipa.v26i3.pp1377-1405
Received: 27 June 2025
Accepted: 20 July 2025
Published: 07 August 2025

suggests that science education has not fully developed students' scientific thinking and contextual problem-solving skills (Black & Wiliam, 2018; Yuberti, Sairi, Nanto, & Sholeha, 2020). Several factors contribute to this problem. First, there is a paradigmatic gap between conventional assessments commonly used in schools and PISA assessment standards. Learning assessments typically focus on memorising concepts rather than measuring students' ability to apply scientific knowledge to solve contextual problems, as expected within the PISA framework (Kruit, Oostdam, van den Berg, & Schuitema, 2018; Rodić, 2018). This is reflected in assessment instruments that are still dominated by questions measuring only Lower-Order Thinking Skills (LOTS), such as remembering and understanding (Nurmalasari & Hertanti, 2021). Second, learning modules and processes do not yet support strengthening students' science literacy abilities. The learning modules used tend to focus solely on content delivery without being balanced with PISA-based practice questions, while the learning process tends to treat science as merely a collection of knowledge that students must memorise (Amarulloh, Utari, & Feranie, 2017; Fonna, Bunawan, & Derlina, 2022). This is related to the lack of teachers' ability to connect physics with real life and other subjects (Pavkov-Hrvojević & Bogdanović, 2019). As a result, students have difficulty understanding physics concepts in depth and fail to connect them to real-world contexts (Löfgren, Weidow, & Enger, 2023). Therefore, the implementation of assessments based on PISA standards is important as a strategic effort to improve students' scientific literacy and scientific problem-solving competencies (She, Stacey, & Schmidt, 2018)

The low science literacy abilities in Indonesia is a problem in education and seriously impacts national economic growth. A lack of scientific literacy hinders individual decision-making and social participation, which ultimately weakens labour productivity and economic growth (Valladares, 2021). Based on data from STEM (2024), only around 18.47% of college students graduate from STEM fields. The study of Sidiq & Permanasari (2022) also revealed that interest in pursuing STEM studies among students aged 13 to 15 remains relatively low. Approximately 217 million secondary school students, or 60% of the world's youth, have not yet reached the minimum reading proficiency level (Vazquez-Lopez & Huerta-Manzanilla, 2021). More than 55% of students do not meet the minimum competencies in literacy and mathematics (Bank, 2020). The misalignment between vocational and higher education curricula and the needs of Industry 4.0 influences this issue (Bank, 2020). Low science proficiency has profound implications, as evidenced by data showing that Indonesia's hoax literacy rate remains moderate (68%), reflecting low critical thinking skills regarding information, leaving Indonesia's digital population struggling to distinguish valid scientific information (Redaksi, 2024). Meanwhile, research by Das, Wibowo, Chui, Agarwal, & Lath (2019) projects that Indonesia will need 7–9 million skilled workers in STEM fields by 2030. This challenge becomes increasingly critical as we approach the Industry 4.0 revolution, where technological adaptation is a key determinant of global competitiveness (Schwab, 2016).

Addressing the science literacy crisis requires a fundamental transformation in the learning assessment system by developing assessment instruments based on the PISA Framework 2025. This latest framework emphasises four key aspects: 1. Science Competencies, comprising 40% ability to explain phenomena scientifically; 30% constructing and evaluating designs for scientific investigations and critically interpreting scientific data and evidence; and 30% researching, evaluating, and using scientific information for decision-making and action; 2. Types of Knowledge range from 38-48% content, 27-33% procedural, and 24-30% epistemic; 3. Context, covering personal (25%), local/national (50%), and global (25%) levels; and 4. Cognitive Level, namely, low, medium, and high (OECD, 2023b). Assessment instruments need to integrate these four dimensions holistically, no longer limited to a partial approach that only measures factual knowledge (Chan & Luo, 2021).

The development of science literacy instruments during 2021-2024 shows various innovations, such as the development of the physics domain (Hijriati, Sahyar, & Derlina, 2021), local wisdom (Hastuti, Setianingsih, & Anjarsari, 2020), as well as the application of the PISA framework at the primary level (Zhang, Liu, & Feng, 2023) and the secondary level (Nasution, Liliawati, & Hasanah, 2019). However, most of these studies still show fragmentation of competencies. Approximately 42,5% of instruments only measure one of the three science literacy competencies. Although local contexts are beginning to be used, they have not yet been systematically integrated into global frameworks such as PISA 2025. This challenge is common in developing countries, which face resource constraints, cultural relevance in assessment development, and a lack of instrument designs that apply real-world student contexts (Johnson, 1999; Ninomiya, 2016). Additionally, Besche-Truthe (2025) noted that many developing countries struggle to meet large-scale assessment standards due to technical capacity and high costs (Besche-Truthe & Seitzer, 2025). This study offers a holistic approach by developing instruments that balance content, procedural, and epistemic dimensions, tailored to the Indonesian context, to measure students' science literacy abilities.

This study aims to develop a valid and reliable science literacy instrument to measure the science literacy abilities of secondary school students. The research questions focus on: 1) how the characteristics of the science literacy instrument developed align with the PISA Framework 2025 and are relevant to the educational context in Indonesia; 2) how is the quality of the developed instrument assessed based on empirical test results; 3) how the profile of students' science literacy abilities is determined based on measurement results using the instrument; and 4) how practical the instrument is to use. The development of the instrument is expected to contribute to improving the quality of science assessment and support efforts to enhance students' science literacy in Indonesia in line with international standards.

▪ METHOD

Participants

This study involved 300 participants, consisting of 292 students and eight teachers. A total of 137 students and four teachers participated in completing the needs analysis questionnaire. For the science literacy instrument trial, participants were selected through purposive sampling. The limited trial was conducted on 30 students, while the field trial involved 125 students. In addition, four teachers provided their perceptions regarding the clarity and difficulty level of the questions. The students involved were 16 years old and under and had studied electricity, waves, and magnetism. They came from A-accredited private schools in the South Tangerang area that had implemented technology such as smartphones or tablets in their learning. The distribution of academic ability was arranged proportionally, namely 30% low, 40% medium, and 30% high.

Research Design and Procedures

This study employed the Research and Development (R&D) method, employing the 4-D development model (Define, Design, Develop, Disseminate) developed by Thiagarajan, Semmel, & Semmel (1974). The 4-D model was chosen because it is suitable for developing educational instruments that require systematic and structured stages.

Define phase, a series of preliminary analyses was conducted to establish the basis for developing relevant and contextual instruments. a) In-depth analysis of the PISA Framework 2025; b) Mapping of scientific literacy competencies based on the PISA 2025 framework; c) Preliminary study of students' current scientific literacy levels through a review of previous research results and field data. Field data was obtained by distributing questionnaires to 137 students and four teachers as a basis for development needs; d) Documentation analysis of practice questions, daily tests, textbooks, the 2013 curriculum, and the independent curriculum as an effort to contextualize the questions with students' real-world environments revealed that most students tended to simply memorize concepts without understanding their application. The define phase was conducted for two months to formulate objectives and instrument development needs thoroughly.

The design phase was carried out for three weeks to design the instrument based on the needs analysis results. It included several important steps, namely: a) preparing an instrument blueprint that refers to the PISA Framework 2025; b) selection of question formats (multiple choice, complex multiple choice and essay); c) preparation of media (pictures, stimuli, tables and graphs); d) determining the distribution of questions based on four main domains, namely: the knowledge domain (content, procedural, and epistemic), the competency domain (explaining phenomena scientifically; designing and evaluating scientific investigation designs; and interpreting scientific data and evidence critically for decision making), the context domain (personal, local/national, and global); and the cognitive level domain (high, medium, low); e) compile a grid of 11 questions based on scientific literacy indicators and competencies; and f) preparing an assessment rubric and scoring guidelines that are in accordance with the competency indicators being measured.

Development, the development phase was conducted over five weeks and included a series of systematic instrument development processes. These included: a) developing an assessment instrument based on a pre-designed design, consisting of 11 questions that refer to the scientific literacy indicators within the PISA framework; b) conducting an instrument validation process by six expert validators in the field of science education and assessment, aimed at assessing the content, construct, and language; c) revising the instrument based on input and suggestions from experts to ensure the quality, clarity, and appropriateness of the questions before use in the implementation phase; d) conducting a limited trial in two schools involving 30 students, this stage aimed to be an initial effort to determine the level of understanding in the learning context; e) revise the instrument based on the results of limited trials of instructions and question structure; f) conduct a broader field trial with 125 students to obtain more representative empirical data for analyzing the validity and reliability of the instrument; g) analyze data based on the results of the scientific literacy instrument trial.

Disseminate, the dissemination stage was carried out for two weeks, which included the publication of scientific articles in nationally accredited journals and the distribution of science literacy instruments to schools that were the research locations.

Instrument

This study employed several data collection techniques involving teachers and students as the primary respondents to obtain comprehensive data in developing a scientific literacy assessment instrument. This study employed a Likert-scale questionnaire to identify perceptions and needs for scientific literacy assessment instruments in schools based on teachers’ and students’ responses. These questionnaire items were structured based on indicators in the PISA Framework 2025. In addition, A document analysis of the evaluation tools used by teachers was conducted, which included: daily test questions, practice questions, questions in textbooks, questions in the 2013 curriculum book, and questions in the independent curriculum module. This study also used a validation sheet to assess the suitability of each question item's content, construction, and language. As a complement, this study also used a questionnaire given to teachers and students to identify their perceptions regarding the clarity and difficulty level of the questions. The categories of student scientific literacy profiles used in this study are presented in Table 1.

Table 1. Categories of students' scientific literacy profiles

Literacy Level	Description
80-100	Very high
66-79	High
56-65	Medium
40-55	Low
0-39	Very Low

(Sudirman, Rusilowati, & Susilaningsih, 2024)

Data analysis

This study employed two data analysis approaches: qualitative and quantitative. Using a descriptive thematic analysis approach, qualitative analysis was used to process feedback from validators, teachers, and students systematically. The analysis process was conducted by thoroughly reading all responses, identifying important recurring statements, and then grouping these statements into main themes such as question clarity, content suitability to indicators, and student understanding. Theme identification was conducted manually and descriptively without going through a formal coding process. Meanwhile, quantitative analysis included: (1) validity testing using the Content Validity Index (CVI) and Content Validity Ratio (CVR); (2) item analysis based on Classical Test Theory (CTT); (3) reliability testing using Omega McDonald and Cronbach's Alpha; and (4) analysis of students' scientific literacy skills based on test scores.

This research instrument has to fulfill the requirements of established quality standards, including validity, reliability, difficulty level, and item discrimination. The criteria used are presented in Table 2.

Table 2. Instrument psychometric criteria

Aspect	Criteria	Description
Content validity	Minimum CVI of 0.80	The instrument has met the content validity requirements based on expert assessment.

Reliability	Omega McDonald and Cronbach's alpha minimum 0.70	The instrument has good internal consistency.
Level of difficulty	Good test items range from 0.30 to 0.80	The questions were neither too easy nor too difficult for the participants.
Discrimination power	Minimum item score	Able to distinguish between high and low ability participants.

(Lawshe, 1975)

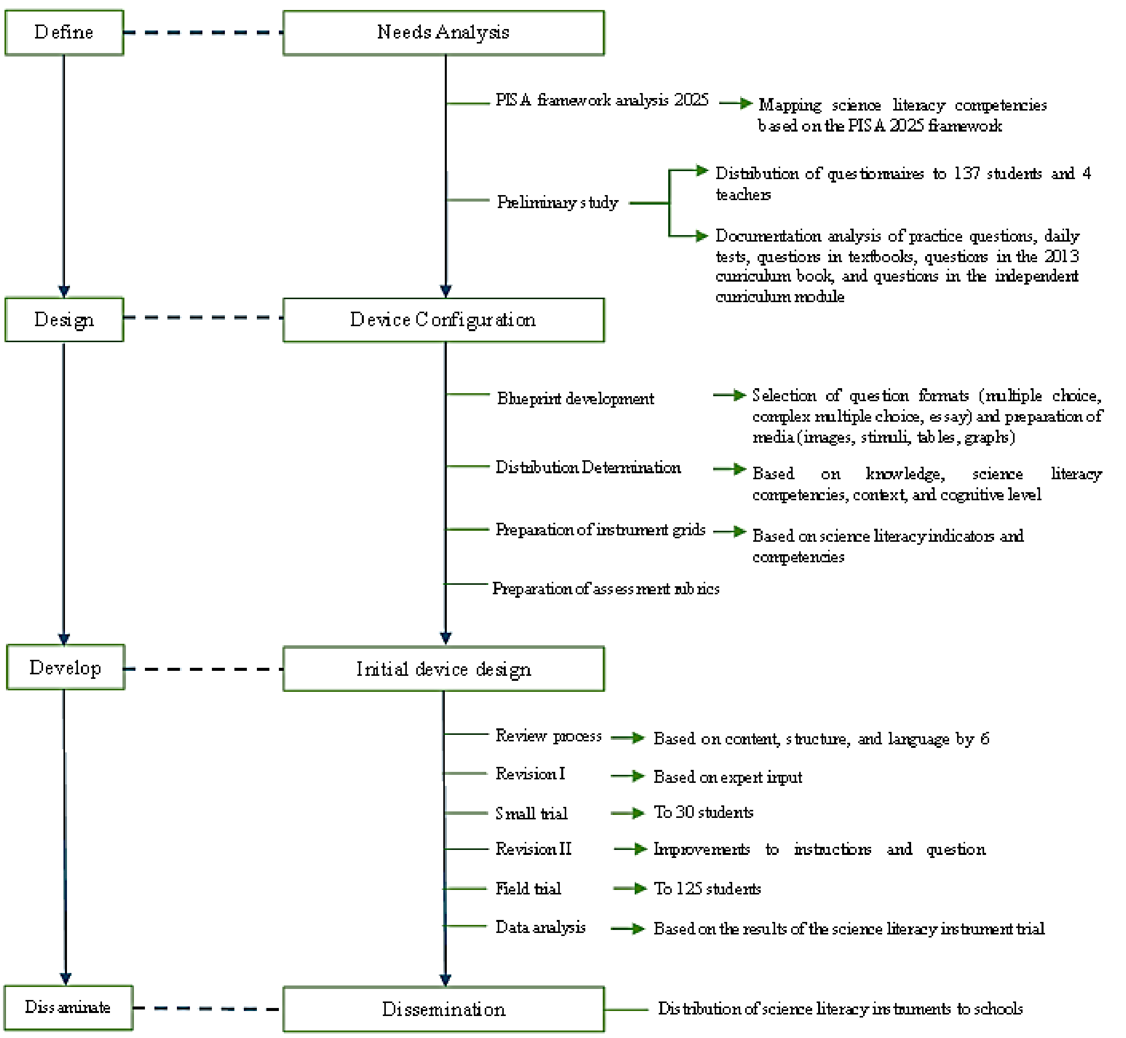

Figure 1. Operational scheme for developing 4D model test instruments

▪ RESULT AND DISSCUSSION

Psychometric Characteristics of Instruments and Their Implications

Preliminary study results indicate that students' science literacy levels are still relatively low, particularly in their ability to explain scientific phenomena scientifically and interpret scientific data and evidence (Amini & Sinaga, 2021). The instruments used by teachers tend to measure low to moderate cognitive levels, requiring only the ability to remember and understand concepts (C1-C2) (Azzopardi & Azzopardi, 2022). Field data also reveal that questions measuring LOTS dominate at 87.2%, while the rest of the questions measure HOTS. Questions related to real-life contexts remain very limited, resulting in underdeveloped critical thinking skills and the application of scientific concepts in everyday situations. These specific findings are directly applied in the design of the instrument framework, focusing on developing items that measure the ability to explain scientific phenomena, interpret data, engage in scientific reasoning, and apply science in real-world contexts. Thus, the instrument framework design is structured based on the specific needs identified from the results of this preliminary study.

Based on the mapping of needs and weaknesses of previous instruments, this study successfully developed a science literacy instrument based on the PISA Framework 2025, consisting of 11 questions covering physics concepts (electricity, waves, magnetism). The instrument is designed to measure three core science literacy competencies with the following distribution: (1) explain scientific phenomena (5 items), (2) design and evaluate designs for scientific investigations and critically interpreting data and scientific evidence (3 items), and (3) researc, evaluate, and use scientific information for decision-making and action (3 items). The question format consists of multiple-choice questions (2 items), complex multiple-choice questions (3 items), and essay questions (6 items) with varying cognitive levels according to the PISA taxonomy. Table 3 presents the distribution of question items.

Table 3. Distribution of question items

No.	Competence	Knowledge	Context	Question Code
1.	Explain scientific phenomena	Content	Personal	L1.1 M1.3 L4.2
			Local	G2.2
			Global	G3.2
2.	Design and evaluate designs for scientific investigations, and critically interpret scientific data and evidence	Procedural	Personal	L4.1
			Local	L2.1
			Global	M3.3
3.	Research, evaluate, and use scientific information for decision making and action	Procedural	Personal	G1.2
		Epistemic	Local	M2.3
			Global	L3.1

The questions in this instrument are presented in various contexts relevant to students' daily lives. For example, there are personal contexts such as wireless charging, local contexts such as earthquakes, and global contexts such as Maglev trains. There are two multiple-choice questions (MCQs), three complex multiple-choice questions (MCQs), and six essay questions. Each question is assigned a specific code, with the letter “G” for wave-related material, “L” for electricity-related material, and “M” for magnetism-related material. The number after the letter indicates the conceptual context: 1 for personal context (wireless charging), 2 for local context (earthquakes), 3 for global context (Maglev trains), and 4 for personal context (experiments). Meanwhile, the number after the dot indicates the question number within that context. For example, the question code L1.1 refers to an electrical material question in a personal context about wireless charging in question number 1.

One hundred fifty-five students were participating in this study. The limited trial involved 30 students, a number generally considered sufficient to assess reliability (Bujang, Omar, Foo, & Hon, 2024). The field trial in this study involved 125 students. This number is considered sufficient for construct validity analysis. Anderson & Gerbing (1984) stated that if each factor is measured with three or more items, then a sample of 100 people is usually sufficient to produce a good model calculation. In addition, Boomsma (1985) also suggests that the sample size should not be less than 100 so that the factor analysis results are more stable. All participants were selected using purposive sampling techniques, considering characteristics relevant to the research objectives. Participant criteria included students aged up to 16 years who had studied electricity, waves, and magnetism and were from schools implementing technology, such as smartphones or tablets, in daily learning activities. The schools involved were private schools with A accreditation located in South Tangerang and surrounding areas. Additionally, the distribution of students' academic abilities was proportional, with 30% in the low category, 40% in the medium category, and 30% in the high category.

Purposive sampling was used to ensure representation of students with varying academic abilities (low, medium, and high), which is an important aspect in evaluating item clarity, readability, and difficulty level. Students aged 16 years or younger were selected to align with the target student ability level measured in PISA, which is generally within that age range. Accredited private schools with an A rating were prioritised because schools with good accreditation are considered to have more structured learning systems and enable optimal implementation of technology-based assessments.

To ensure that the science literacy instruments developed meet psychometric quality standards, a comprehensive evaluation covering content validity, item validity, reliability, and analysis of difficulty and discriminating power is conducted. The following results show how the three aspects influence each other.

The validation of the science literacy instrument involved five lecturers and one secondary school teacher. The validators consisted of experts in physics, science literacy instrument evaluation, and language. The validation process was conducted using an assessment sheet covering three main domains: (1) content, (2) construct, and (3) language. These validation aspects can be seen in Tables 4 and 5. Each domain was evaluated based on specific aspects tailored to the question format, namely multiple choice, complex multiple choice, and essay questions. The assessment was conducted using the Guttman scale and included a comments column for feedback or suggestions for improvement. The presence of science literacy instrument experts and physics subject matter experts aimed to ensure that the questions were not only aligned with the PISA 2025 framework but also conceptually and contextually accurate.

Table 4. Validation aspects of multiple-choice science literacy instruments

Context	Rated Aspect
Content	1. Suitability of questions with learning indicators
	2. Logical answers/no misconceptions
	3. Suitability in everyday life
	4. Suitability with aspects of scientific literacy
Construct	1. The main question is formulated clearly
	2. The question does not provide clues to the answer

	3. Images/tables and the like are clear and functional
	4. There are scoring guidelines
	5. The main question does not use double negative sentences
Language	1. Compliance with Indonesian language rules
	2. Using communicative language
	3. Using common language/words (not regional languages)
	4. Sentence formulation does not cause misunderstanding

Table 5. Validation aspects of complex multiple choice and essay science literacy

Context	Rated Aspect
Content	1. Suitability of questions with learning indicators
	2. Logical answers/no misconceptions
	3. Suitability in everyday life
	4. Suitability with aspects of scientific literacy
Construct	1. The main question is formulated clearly
	2. The question does not provide clues to the answer
	3. Images/tables and the like are clear and functional
	4. There are scoring guidelines
	5. Using question words and commands that require descriptive answers
Language	1. Compliance with Indonesian language rules
	2. Using communicative language
	3. Using common language/words (not regional languages)
	4. Sentence formulation does not cause misunderstanding

After the validation process was carried out qualitatively by experts, the instrument's content validity was then analysed quantitatively using the Content Validity Index (CVI) and Content Validity Ratio (CVR) indices based on the theory (Lawshe, 1975). The results of the analysis in Table 6 show that the instrument meets the ‘Highly Valid’ criteria in all three domains, with the following details:

Table 6. Expert validation results of science literacy instrument (CVR)

Question Code	CVR		
	Material	Construct	Language
G1.2	1.00	0.90	0.90
M1.3	1.00	1.00	1.00
G2.2	0.90	1.00	1.00
M2.3	0.80	0.90	0.90
G3.2	0.90	1.00	0.90
L1.1	0.80	1.00	0.90
L2.1	0.80	1.00	1.00
L3.1	0.80	0.90	0.90
M3.3	0.80	1.00	0.80
L4.1	1.00	1.00	0.90
L4.2	1.00	1.00	0.90

Table 7. Expert validation results of the science literacy instrument (CVI)

Aspect	CVI	Criteria
Material	0.90	Highly Valid

Construct	1.00	Highly Valid
Language	0.90	Highly Valid

CVI Value: 0.93 (Highly Valid)

The CVR analysis results showed values ranging from 0.8 to 1.0, higher than the critical value of 0.78 for six validators (Lawshe, 1975). A CVI value of 0.93 indicates 'highly valid' content validity, based on the criteria (Polit & Beck, 2006) requiring CVI ≥ 0.90 for high-quality instruments. Interestingly, the construct aspect achieved a perfect CVI (1.00), indicating the success of the instrument development in representing the PISA 2025 framework. This finding aligns with the research of Retnawati (2016) stated that the instruments based on international frameworks tend to have high validity when systematically adapted to local contexts (Borup, Shin, Powell, Evmenova, & Kim, 2022). The high CVR and CVI values (Azwar, 2013) not only ensure content validity but also indicate the adequacy of representation of the science literacy construct domains being measured (Suhaini, Ahmad, & Bohari, 2021).

After ensuring the validity of the content through CVR and CVI analysis, the quality of items was analyzed. This study uses the Classical Test Theory (CTT) approach to calculate item validity, reliability, difficulty level, and item discrimination. The CTT approach was chosen because the number of respondents in this study was limited, namely 125 participants. Meanwhile, Item Response Theory (IRT) requires a larger sample size, typically ranging from 200 to 500 respondents, especially for more complex models (Nguyen, Han, Kim, & Chan, 2014). Nevertheless, using CTT still allows researchers to obtain sufficient information about the quality of the test items.

Based on the analysis of item validity using SPSS, the results show that essay and complex multiple-choice questions have an average validity of 0.55. In contrast, multiple-choice questions have an average validity of 0.87. The overall average validity of the instrument is 0.71, indicating that the items are generally valid and capable of consistently measuring students' science literacy skills. This aligns with the guideline that correlation values above 0.70 are categorized as 'Highly Valid' (Arikunto, 2018).

However, some items showed moderate to low validity values, such as M1.3 (0.32) and M3.3 (0.28), indicating the need for revision. Qualitative analysis shows that the validity of these items is not optimal due to the high cognitive demands that require analytical skills, data synthesis, and decision-making based on scientific context. For example, in the Maglev system question (M3.3), students are confronted with technical terminology such as EDS, LSM, and LSRM, accompanied by complex narrative information, making it difficult to understand and select the appropriate answer. The validity of questions depends heavily on the clarity of wording, the relevance of the context to students' experiences, and the proportional level of difficulty (Guillot-Valdés, Guillén-Riquelme, & Buela-Casal, 2022; Pluye et al., 2014). Therefore, improvements to these items focused on simplifying the language, presenting contexts closer to students' daily lives, and strengthening alternative, more logically equivalent answers.

In addition to the construction of the questions, the low validity of some questions was also due to student characteristics. For example, in questions M3.1, M3.3, and L4.2, which had a validity of less than 0,40, many students did not complete the questions because they were not accustomed to solving questions with a global context or complex experimental data. Four out of five schools have educational backgrounds that do not fully support a science literacy approach, resulting in students struggling to connect information across paragraphs, graphs, and tables.

Reliability analysis was conducted using two different approaches, tailored to the questions developed. Table 8 presents the results of reliability analysis using McDonald's Omega coefficient for nine questions (essays and complex multiple choice) and Cronbach's Alpha for two questions (multiple choice) based on field trials.

Table 8. Results of the items reliability test in the field test

Type of question	Omega MCDonald	N of items
Essay and complex multiple-choice	0.79	9
Type of question	**Cronbach's Alpha**	**N of items**
Multiple choice	0.69	2

The reliability analysis revealed interesting differences between question formats. For essay questions and complex multiple-choice questions analysed together, the reliability coefficient was calculated using Omega McDonald and yielded a value of 0.79, which falls into the 'good' category ($\omega > 0.70$) and is considered adequate for the development of new instruments (Viladrich, Angulo-Brunet, & Doval, 2017). Meanwhile, regular multiple-choice questions consisting of only two items were analysed separately using Cronbach's Alpha and obtained a value of 0.69, which is at the lower limit of the 'adequate' category (Taber, 2018). The limited number of multiple-choice questions (n=2) contributed to the moderate reliability. Simulations indicate that increasing the number of multiple-choice questions to 6 could potentially raise α to 0.82, meeting the 'good' standard for high-stakes assessment (Nunnally & Bernstein, 1994).

The results of Pearson's correlation analysis between G1.2 and G2.2 showed a value of $r = 0.52$ ($p<0.70$), which falls into the moderate relationship category because $0.50 \leq r < 0.70$. This indicates that the two multiple-choice questions measure related aspects of the science literacy construct. The reliability for both multiple-choice questions (G1.2 and G2.2) was reported separately using Pearson's correlation between items ($r=0.52; p<0.001$). This value is indicative because reliability cannot be calculated using the omega coefficient due to the small number of items (two items). Therefore, the multiple-choice questions were not combined in the omega analysis for the combined complex multiple-choice and essay questions.

The selection of different reliability approaches in this study was based on technical and theoretical reasons. Complex multiple-choice questions and essay questions were combined because they measure the same construct, as verified through construct validation (Rodriguez, 2003). Meanwhile, multiple-choice questions were not combined because they have different constructs. The omega coefficient was used to calculate the reliability of the combined scores (complex multiple-choice and essay questions) because it is more robust in handling multidimensional data and does not require the tau-equivalence assumption like the alpha coefficient (McDonald, 1999). Multiple-choice questions were not included in the omega calculation because there were only two items, thus failing to meet the minimum requirement for analysis in SPSS version 31. If there are at least three multiple-choice questions with the same construct, these questions can be combined with complex multiple-choice and essay questions for data analysis. The difficulty level of the questions in this study was classified based on the percentage of students who answered correctly. The difficulty categories are presented in Table 9.

Table 9. Level of difficulty of questions

Type of Question	Question Code	Difficulty Index (p)
Essays and Complex Multiple Choice	L1.1	0.37
	M1.3	0.45
	L2.1	0.37
	M2.3	0.24
	L3.1	0.36
	G3.2	0.36
	M3.3	0.11
	L4.1	0.46
	L4.2	0.19
Multiple Choice	G1.1	0.63
	G2.2	0.54

The level of difficulty of questions can be interpreted based on the following criteria: if $p>0.70$, then the question is classified as easy; if $0.30 \leq p \geq 0.70$, the question is classified as moderate; and if $p<0,30$, then the question is classified as difficult. Based on the analysis of eleven questions, consisting of two multiple-choice questions and nine essay and complex multiple-choice questions, eight questions (72.73%) were in the moderate category, while three questions (27.27%) were in the difficult category. Generally, the fewer students who answer a question, the higher the difficulty level of the question, and conversely, the more students who answer correctly, the lower the difficulty level tends to be (Iñarrairaegui et al., 2022).

These results regarding the level of difficulty are reinforced by data from student response questionnaires after completing the science literacy questions. As many as 77.60% of students stated that they had never completed science literacy questions, particularly those based on PISA. Of the difficulty level, 22.40% of students rated the questions as very difficult, 34,40% categorised them as difficult, and 28.80% stated that the difficulty level was moderate. These findings align with research of Alatlı (2020) and Le Hebel, Montpied, Tiberghien, & Fontanieu (2017), indicating that students struggle with PISA-based science literacy questions due to unfamiliarity with the question format.

The results of the difficulty level analysis indicate that not all questions are at the same theoretical and empirical difficulty levels. Most questions theoretically categorized as ‘easy’ are empirically classified as “moderate” or even ‘difficult’. For example, questions L1.1 and M1.3 were designed at a low cognitive level (easy), but empirical test results showed a ‘moderate’ difficulty level with correct answer proportions of 0.37 and 0.45, respectively. Similarly, question G3.2 was theoretically easy, but empirically also fell into the moderate difficulty level (0.36). Meanwhile, there is also question M3.3, which is theoretically at a moderate difficulty level but empirically proven to be ‘difficult’ (correct response rate of 0.11). These results indicate inconsistencies between the theoretical cognitive level and empirical difficulty. Some questions were designed to be easy but turned out to be quite challenging for students, while some questions with higher cognitive levels were answered with better success rates. This confirms that, in addition to cognitive levels, factors such as stimulus clarity, familiar question contexts, and language style also influence the empirical difficulty level of a question.

It is essential to consider other factors that influence these results. Questions with long stimuli, such as those in PISA, may not only test science literacy but also students' reading literacy skills. Limitations in understanding complex texts can lead to failure in answering easy questions (Klotz, Ehmke, & Leiss, 2025; Pongsakdi et al., 2020). Previous needs analysis results indicate that learning that focuses on memorisation rather than scientific reasoning can hinder students' readiness to solve questions requiring understanding and applying concepts.

Overall, students agreed that the questions required critical thinking skills (HOTS), which not only rely on memory but also require analytical, critical, and creative thinking skills to understand and answer the questions correctly (Suwarna & Fatimah, 2018). Interviews with 15 students as a validation sample showed consistent responses to the questionnaire data. A study by OECD (2019) explained that the characteristics of PISA questions were designed to measure students' analytical and evaluation skills in real-life contexts.

Further analysis was conducted on the discriminatory power of the questions to determine the ability of each item to differentiate between groups of students with high and low scientific literacy. Table 10 presents the results of the discriminatory power analysis based on the field test.

Table 10. Results of the analysis of the discriminatory power of question items in the field Test

Type of question	Question code	Discriminatory power index (D)
Essays and complex multiple-choice	L1.1	0.41
	M1.3	0.25
	L2.1	0.64
	M2.3	0.60
	L3.1	0.30
	G3.2	0.72
	M3.3	0.17
	L4.1	0.55
	L4.2	0.29
Multiple choice	G1.1	0.52
	G2.2	0.52

Interpretation of the discriminatory power coefficient can use the criteria developed by Ebel as follows: an index < 0.20 is classified as poor, an index of 0.20-0.40 as satisfactory, an index of 0.40-0.70 as good, and an index of 0.70-1.00 as excellent (Ropii & Fahrurrozi, 2017). Based on the results of the discriminatory power analysis in Table 10, the scientific literacy test instrument with essay questions and complex multiple-choice items shows the following distribution of discriminatory power: very good, 11.11%; good, 44.44%; sufficient, 33.33%; and less than 11.11%. Meanwhile, all multiple-choice items are 100% included in the good criteria. Overall, when combined with essay questions, complex multiple choice, and multiple choice, the distribution of the discriminatory power of the scientific literacy test instrument is very good, 9.09%;

good, 54.54%; sufficient, 27.27%; and less than 9.09%. These results show that 63.63% of the questions are classified as very good and good, so that the test instrument can differentiate between high and low ability students.

As many as 36.36% of the questions were still in the fair and poor categories, including questions L4.2, M1.3, and M3.3, which showed low surgical power. The low discriminating power in these three questions indicates that students experience conceptual and procedural misconceptions. Question L4.2 requires students to conduct an interactive simulation to investigate the relationship between metal type and electric current, then interpret the resulting data in tabular form. The main obstacle lies in students' unfamiliarity with using digital simulations and limited skills in reading and organising experimental data scientifically.

Meanwhile, question M1.3 asks students to identify factors influencing energy transfer efficiency based on two experimental tables. The low discriminating power of this question is due to students' weak ability to distinguish between independent, dependent, and control variables, as well as difficulty understanding the quantitative relationship between experimental parameters and efficiency. Furthermore, the complex multiple-choice question format, requiring two correct answers, potentially adds to the challenge, especially for students unfamiliar with similar question formats (Le Hebel, Tiberghien, Montpied, & Fontanieu, 2019). Similarly, question M3.3 requires decision-making based on technical data to design a Maglev train system. The high complexity of this question relates to the integration of narrative information and comparative tables, as well as the context of advanced technology that is relatively unfamiliar to most students. Therefore, improvements to the question are insufficient through editorial revision alone; they also require considering students' conceptual readiness and scientific thinking skills in a real-world context.

Time constraints also influenced the results of the discriminating power analysis. When working on the science literacy questions, students were given 80-90 minutes (equivalent to two class hours). During the implementation, 95% of students reported not completing the 11 questions optimally. This complaint was reinforced by the fact that only 5% of students answered all the questions on time. This reflects limited time management and reveals gaps in students' critical thinking skills and conceptual understanding of science literacy, particularly on questions based on real-world contexts such as the PISA model. Questions with low discriminating power are likely influenced by time constraints, where students do not have sufficient opportunity to answer thoroughly, resulting in suboptimal results (Lu & Sireci, 2007). These results are consistent with research by Shaffer, Ferguson, & Denaro (2019), which states that science literacy questions generally have long texts and require critical thinking, thus requiring more time to understand the reading and questions. Further observations during the implementation showed that students tended to spend more time on question L4.1. This aligns with the characteristics of PISA scientific literacy, which emphasises the inquiry process (OECD, 2019), where HOTS acts as a determining factor in the speed of problem solving. In addition to improving the quality of questions based on their discriminating power, it is also necessary to consider adjusting time allocation to allow students to work more optimally on the questions.

Portrait of Indonesian Students' Scientific Literacy: A Reflection on Science Education

Efforts to improve the quality of science learning in Indonesia must begin with understanding the actual conditions on the ground. The students' scientific literacy achievements in Figures 2 and 3 present a reality that deserves our attention. This data visualisation can serve as a starting point for developing instruments and designing more contextual learning strategies.

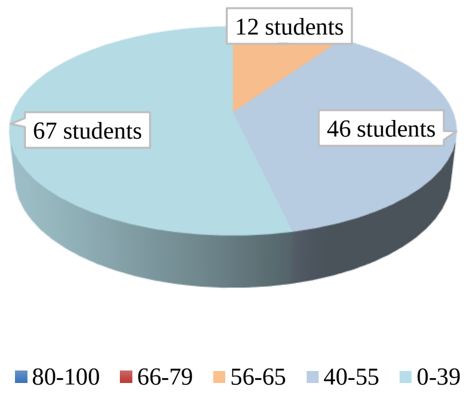

Figure 2. Results of students' scientific literacy scores

The results of the analysis show that students' overall scientific literacy remains very low. As many as 53.60% of students fall into the very low category, 36.80% into the low category, and only 9.60% reach the moderate level. None of the students got the high or very high category. These findings reflect the low level of scientific literacy among students and are a serious indicator of the quality of current science learning.

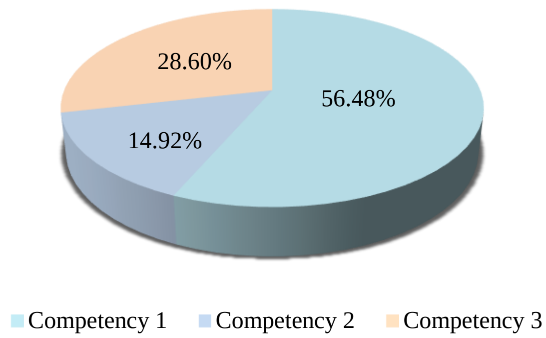

Figure 3. Results of students' scientific literacy based on competencies

The disparity in achievement is evident when examining the differences between competencies. The highest achievement was in the competency of "explaining scientific phenomena" (56.48%), while inquiry competencies such as "evaluating scientific investigation designs" (14.92%) and "using scientific information for decision-making" (28.60%) were still low. This indicates significant weaknesses in critical thinking and scientific investigation skills. This pattern aligns with the results of the 2022 PISA test, where only 34% of Indonesian students achieved Level 2 (basic ability to explain phenomena), and less than 1% mastered Levels 5-6 (complex inquiry competencies). When compared with neighbouring countries, these weaknesses appear systemic across the region, albeit with significant variation. For example, Singapore demonstrated strong mastery, with 92% of students achieving Level 2 and 24% at Levels 5-6. Meanwhile, although higher than Indonesia, Malaysia (52% Level 2, 1% Level 5-6) is still below the OECD average (76% Level 2, 7% Level 5-6). This finding indicates systematic weaknesses in inquiry-based science learning (She et al., 2018), both at the local and national levels. Therefore, fundamental improvements in the science curriculum and learning methodology are needed for Indonesia to catch up with other countries in the region.

Furthermore, based on the data obtained, the percentage of student achievement in scientific literacy can be categorised based on three main contexts: personal (60.21%), local (27.35%), and global (12.44%). These results indicate that students can apply scientific literacy in a personal context, while understanding in local and global contexts is relatively lower. A high percentage in the personal context indicates that students find it easier to understand and apply scientific concepts directly related to everyday life or individual issues. Lower achievement in local and global contexts indicates that students have difficulty connecting science to broader issues at the local and global levels.

Students' understanding of each topic varied significantly. For electricity, 40.32% of students demonstrated good and competent skills. Meanwhile, only 32.68% of students demonstrated good scientific literacy skills for waves, and the percentage dropped to 27.00% for magnetism. These data indicate that students' scientific literacy tends to be stronger for electricity. The weak understanding of waves and magnetism is thought to be related to inequalities in the 2013 curriculum, which prioritised science (basic physics and mathematics) as the primary focus, while technology or applied concepts such as waves and magnetism received less emphasis in the content structure of textbooks (Herlanti, Amalia, & Nurlaela, 2022). Therefore, efforts are needed to emphasise and strengthen learning, particularly through integrating applied contexts and in-depth exploration of waves and magnetism.

The main obstacles experienced by students lie in understanding long narrative stimuli and their ability to use virtual experiments. According to Bybee & McCrae (2009), this is caused by students' lack of ability to understand long stimuli, inaccuracy in reading instructions, and limited experience conducting virtual experiments (Linn, 2003). Similar findings were reported by Gok & Goldstone (2024) and Teig (2020), who revealed that students had difficulty analysing data patterns, drawing conclusions from simulation graphs, and connecting virtual experimental data to scientific concepts. This condition is reflected in students' performance answering question M3.3, where most students did not provide correct answers. This indicates that students have not been trained in designing and evaluating scientific investigation designs and are not yet accustomed to critically interpreting scientific data or evidence (Stein, Smith, & Holmes, 2019; Walsh, Quinn, & Holmes, 2019).

Referring to Table 10, question M3.3 is considered difficult because only a few students can answer it. Analysis of students' incorrect answers revealed that the main errors were caused by inaccurate reading of the question, resulting in students incorrectly identifying the solution steps and a lack of understanding of the stimuli provided, such as tables and supporting text, resulting in an inability to connect the information to the concepts being tested. These errors indicate weak reading literacy and scientific information processing skills in students. Furthermore, students are also noted to lack training in explaining scientific phenomena, as seen in question L4.2, which examines students' understanding of the phenomenon of metal resistivity to electric current. Analysis of incorrect answers to question L4.2 revealed a pattern of interpretive errors, where students struggled to connect physics concepts to the context of the question. Most incorrect answers indicated misconceptions about the relationship between resistivity and electric current. Similar difficulties were found in question M2.3, where students struggled to answer questions that assess their ability to research, evaluate, and use scientific information to make decisions. The error patterns in this question were procedural and interpretive, such as an inability to read tables correctly and failing to extract relevant information from the stimuli.

This finding is consistent with the study by Purwaningsih, Sari, & Suryadi (2020), which revealed that evaluation skills remain a challenge. This is evident from the decline in scores in the answer assessment stage compared to previous stages, such as identification and solution planning, which is caused by students' lack of habituation in developing critical evaluation competencies. The few correct answers on the three questions indicate that students' scientific literacy in Indonesia is still relatively low, considering that the three questions represent all the main competencies in scientific literacy.

Practicality of the Instrument

The instrument's practicality test involved four respondents: one science teacher, two physics teachers, and one vice principal for curriculum, representing seven secondary schools in South Tangerang, the trial location. The test results showed that the instrument met the practicality criteria with an average score of 78.85%. Specifically, the test instructions scored 70% (practical category), time effectiveness 75% (practical), relevance to learning 89.20% (very practical), and ease of use 81.20% (practical) (Samsudin, Sadiman, & Pachrozi, 2019). The involvement of the curriculum representatives aimed to assess the instrument's suitability to the school's learning structure. According to the teachers, the ideal time to complete the science literacy questions ranges from 30 to 60 minutes. In general, the teachers showed a positive appreciation for the science literacy test instrument that had been developed. Teachers considered the questions quite reasonable, interesting, creative, and appropriate for giving to students. Several teachers also provided development suggestions: the need to expand its application to other science subjects and improve the practical questions to make the instructions more detailed and easier for students to understand. In line with this, Arikunto also believes that a practical test is a test that is easy to administer, easy to check, and equipped with clear instructions for use (Arikunto, 2018).

Research Limitations

This study shows a paradox in the scientific literacy assessment system in Indonesia. Although the developed instrument has met the validity standards (CVI = 0.93), reliability of essays and complex PG (ω = 0.79) and PG (α = 0.69), several limitations need to be acknowledged: 1. the relatively moderate reliability value is influenced by the limited number of multiple-choice questions (only two questions) and the unbalanced composition of questions with other types (essays/complex PG), 2. purposive sampling from one geographic area limits the generalizability of the findings to more diverse national contexts, 3. the Classical Test Theory (CTT) approach used has weaknesses compared to Item Response Theory (IRT), such as the inability to identify item bias in more depth, and 4. the risk of 'teaching to the best' where this instrument has the potential to be used only to train students to answer PISA-style questions, instead of encouraging authentic, problem-solving-based science learning.

Based on the identified limitations, further research needs to focus on refining and utilising the instrument more holistically. 1. The number and variety of questions, particularly multiple-choice questions, should be increased with a balanced composition of essay or complex PG questions to improve reliability and construct validity. 2. The scope of the sample needs to be expanded through a stratified random sampling method involving diverse geographic regions and student ability levels, so that the findings can be generalized to the heterogeneous Indonesian context. 3. The Item Response Theory (IRT) approach is also essential to analyse question bias more rigorously, while simultaneously measuring individual student abilities. 4. On the practical side, developing an instruction module for teachers is needed, including guidance on using the instrument as a diagnostic tool and examples of authentic science literacy-based learning to prevent the practice of 'teaching to the test'. Dissemination activities should be expanded through journal publications and tiered teacher training, collaboration with policymakers, and the open provision of the instrument for widespread adaptation. Implementing these recommendations is expected to improve the quality of science literacy assessments while encouraging systematic improvements in science learning.

▪ CONCLUSION

The scientific literacy test instrument, developed based on the PISA Framework 2025, consists of eleven questions covering physics topics (electricity, waves, and magnetism) and integrating aspects of scientific literacy. The analysis shows that this instrument has an average validity of 0.71, which is considered valid, and reliability with an Omega coefficient ($\omega = 0.79$) and Cronbach's Alpha ($\alpha = 0.69$) that is included in the good category. In terms of difficulty level, the questions cover a variety of levels from moderate to difficult, while the average discriminatory power indicates effectiveness in differentiating student abilities. Analysis of the test results reveals that students' scientific literacy is generally still very low, reflecting significant challenges in science learning in Indonesia. This aligns with the downward trend in Indonesia's PISA score (383 in 2022) and reinforces the urgency of science education reform. This low literacy has broader impacts, including a STEM skills deficit, a potential decline in GDP of 1.5–2.5%, and a lack of critical thinking skills in responding to scientific information. Overall, this instrument is suitable for use as a scientific literacy measure (with a practicality of 78.85%) and has value as a diagnostic tool to identify systemic weaknesses in the science learning process. This makes it relevant to support the transformation of PISA-based education assessments into a holistic approach by 2025. Therefore, these findings emphasize the need for evidence-based interventions in curriculum, pedagogy, and teacher training and support using this instrument as a basis for formulating more targeted and sustainable scientific literacy policies.

▪ REFERENCES

Alatlı, B. (2020). Cross-cultural measurement invariance of the items in the science literacy test in the programme for international student assessment (PISA-2015). *International Journal of Education and Literacy Studies*, *8*(2), 16. https://doi.org/10.7575/aiac.ijels.v.8n.2p.16

Amarulloh, R. R., Utari, S., & Feranie, S. (2017). The implementation of levels of inquiry with writing-to-learn assignment to improve vocational school students' science literacy. *Journal of Physics: Conference Series*, *812*(1), 012049. https://doi.org/10.1088/1742-6596/812/1/012049

Amini, S., & Sinaga, P. (2021). Inventory of scientific literacy ability of junior high school students based on the PISA framework competency criteria *Journal of Physics: Conference Series*, *1806*(1), 012017. https://doi.org/10.1088/1742-6596/1806/1/012017

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *PSYCHOMETRIKA*, *49*(2), 155–173.

Arikunto, S. (2018). *Dasar-Dasar evaluasi pendidikan [fundamentals of educational evaluation]* (3rd ed.; R. Damayanti, Ed.). Jakarta: PT Bumi Aksara. Retrieved from https://books.google.co.id/books?id=j5EmEAAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false

Azwar, S. (2013). *Penyusunan skala psikologi [development of psychological scales]* (II). Yogyakarta: Pustaka Pelajar.

Azzopardi, M., & Azzopardi, C. (2022). Biology essay questions: cognitive level as revealed by bloom's taxonomy. *PUPIL: International Journal of Teaching, Education and Learning*, *5*(3), 99–111. https://doi.org/10.20319/pijtel.2022.53.99111

Bank, W. (2020). *Janji Pendidikan Indonesia* [indonesia's education promise].

Besche-Truthe, F., & Seitzer, H. (2025). Testing for the money: an analysis on the interdependence of participation in international large-scale assessments and development aid networks. *Frontiers in Education*, *10*(April), 1–13. https://doi.org/10.3389/feduc.2025.1429107

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *PSYCHOMETRIKA*, *50*(2), 229–242.

Borup, J., Shin, J. K., Powell, M. G., Evmenova, A. S., & Kim, W. (2022). Revising and validating the community of inquiry instrument for moocs and other global online courses. *The International Review of Research in Open and Distributed Learning*, *23*(3), 82–103. https://doi.org/10.19173/irrodl.v23i2.6034

Bujang, M. A., Omar, E. D., Foo, D. H. P., & Hon, Y. K. (2024). Sample size determination for conducting a pilot study to assess reliability of a questionnaire. *Restorative Dentistry and Endodontics*, *49*(1), 1–8. https://doi.org/10.5395/rde.2024.49.e3

Bybee, R. W., & McCrae, B. (2009). *PISA Science 2006: Implications for science teachers and teaching*. National Science Teachers Association.

https://doi.org/10.2505/9781933531311

Chan, C. K. Y., & Luo, J. (2021). A four-dimensional conceptual framework for student assessment literacy in holistic competency development. *Assessment & Evaluation in Higher Education*, *46*(3), 451–466. https://doi.org/10.1080/02602938.2020.1777388

Costa, A., Loureiro, M., & Ferreira, M. E. (2021). Scientific literacy: the conceptual framework prevailing over the first decade of the twenty-first century. *Revista Colombiana de Educación*, *1*(81), 195–228. https://doi.org/10.17227/rce.num81-10293

Das, K., Wibowo, P., Chui, M., Agarwal, V., & Lath, V. (2019). Automation and the future of work in Indonesia. *McKinsey & Company* (September). Retrieved from https://www.mckinsey.com/featured-insights/asia-pacific/automation-and-the-future-of-work-in-indonesia

Elhai, J. (2023). Science literacy: a more fundamental meaning. *Journal of Microbiology & Biology Education*, *24*(1), 10–16. https://doi.org/10.1128/jmbe.00212-22

Fonna, N., Bunawan, W., & Derlina. (2022). Development of teaching materials like PISA for the physics mechanical wave topic in high school. *Journal of Physics: Conference Series*, *2193*(1), 012065. https://doi.org/10.1088/1742-6596/2193/1/012065

Gok, S., & Goldstone, R. L. (2024). How do students reason about statistical sampling with computer simulations? An integrative review from a grounded cognition perspective. *Cognitive Research: Principles and Implications*, *9*(1). https://doi.org/10.1186/s41235-024-00561-x

Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2022). Content validity through expert judgment for the depression clinical evaluation test. *International Journal of Clinical and Health Psychology*, *22*(2). https://doi.org/10.1016/j.ijchp.2022.100292

Hastuti, P. ., Setianingsih, W., & Anjarsari, P. (2020). How to develop students ' scientific literacy through integration of local wisdom in Yogyakarta on science learning ? *Journal of Physics: Conference Series*, *1440*, 1–7. https://doi.org/10.1088/1742-6596/1440/1/012108

Herlanti, Y., Amalia, U., & Nurlaela, A. (2022). Profile of school readiness in applying stem (science, technology, engineering, mathematics) education. *Edusains*, *14*(1), 14–23. https://doi.org/10.15408/es.v14i1.25541

Hijriati, Sahyar, & Derlina. (2021). The development of physics test instrument based on pisa for optical topic in high school. *Journal of Physics: Conference Series*, *1811*(1), 012039. https://doi.org/10.1088/1742-6596/1811/1/012039

Iñarrairaegui, M., Fernández-Ros, N., Lucena, F., Landecho, M. F., García, N., Quiroga, J., & Herrero, J. I. (2022). Evaluation of the quality of multiple-choice questions according to the students' academic level. *BMC Medical Education*, *22*(1), 779. https://doi.org/10.1186/s12909-022-03844-3

Johnson, S. (1999). International association for the evaluation of educational achievement science assessment in developing countries. *Assessment in Education: Principles, Policy & Practice*, *6*(1), 57–73. https://doi.org/10.1080/09695949992991

Klotz, E., Ehmke, T., & Leiss, D. (2025). Text comprehension as a mediator in solving mathematical reality-based tasks: the impact of linguistic complexity, cognitive factors, and social background. *European Journal of Educational Research*, *14*(1), 23–39. https://doi.org/10.12973/eu-jer.14.1.23

Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Assessing students' ability in performing scientific inquiry: instruments for measuring science skills in primary education. *Research in Science & Technological Education*, *36*(4), 1–27. https://doi.org/10.1080/02635143.2017.1421530

Lawshe, C. H. (1975). A quantitative approach to content validity 1. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Le Hebel, F., Montpied, P., Tiberghien, A., & Fontanieu, V. (2017). Sources of difficulty in assessment: example of PISA science items. *International Journal of Science Education*, *39*(4), 468–487. https://doi.org/10.1080/09500693.2017.1294784

Le Hebel, F., Tiberghien, A., Montpied, P., & Fontanieu, V. (2019). Teacher prediction of student difficulties while solving a science inquiry task: example of PISA science items. *International Journal of Science Education*, *41*(11), 1517–1540. https://doi.org/10.1080/09500693.2019.1615150

Linn, M. (2003). Technology and science education: Starting points, research programs, and trends. *International Journal of Science Education*, *25*(6), 727–758. https://doi.org/10.1080/09500690305017

Löfgren, S. K., Weidow, J., & Enger, J. (2023). Rolling balls or trapping ions? How students relate models to real-world phenomena in the physics laboratory. *Science Education*, *107*(5), 1215–1237. https://doi.org/10.1002/sce.21802

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29–37. https://doi.org/10.1111/j.1745-3992.2007.00106.x

McDonald, R. P. (1999). *Test Theory: A Unified Treatment* (1st ed.). Mahwah: Lawrence erlbaum associates. https://doi.org/10.2307/2669496

Nasution, I. B., Liliawati, W., & Hasanah, L. (2019). Development of scientific literacy instruments based on pisa framework for high school students on global warming topic. *Journal of Physics: Conference Series*, *1157*(3). https://doi.org/10.1088/1742-6596/1157/3/032063

Nguyen, T. H., Han, H.-R., Kim, M. T., & Chan, K. S. (2014). An introduction to item response theory for patient-reported outcome measurement. *The Patient*, *7*(1), 23–35. https://doi.org/10.1007/s40271-013-0041-0

Ninomiya, S. (2016). The impact of PISA and the interrelation and development of assessment policy and assessment theory in Japan. *Assessment in Education: Principles, Policy & Practice*, *26*(1), 91–110. https://doi.org/10.1080/0969594X.2016.1261795

Nisak, F., & Yulkifli, Y. (2021). Development of electronic module using inquiry based learning (IBL) model integrated high order thinking skill (HOTS) in 21 st century physics learning class X. *Journal of Physics: Conference Series*, *1876*(1), 012085. https://doi.org/10.1088/1742-6596/1876/1/012085

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory, masalah 972*. New York: McGraw-Hill.

Nurmalasari, M., & Hertanti, E. (2021). The effect of guided inquiry based hypermedia on students' high order thinking skills in thermodynamics concepts. *Journal of*

Physics: Conference Series, *1836*(1), 012062. https://doi.org/10.1088/1742-6596/1836/1/012062

OECD. (2019). *PISA 2018 results (volume I)*. OECD. https://doi.org/10.1787/5f07c754-en

OECD. (2023a). *PISA 2022 Results (Volume I)*. Paris: OECD. https://doi.org/10.1787/53f23881-en

OECD. (2023b). *PISA 2025 science framework (drAFT)*. OECD Publishing. Retrieved from https://pisa-framework.oecd.org/science-2025/

Pavkov-Hrvojević, M., & Bogdanović, I. (2019). Making real-life connections and connections between physics and other subjects. *AIP Conf*, 180013. https://doi.org/10.1063/1.5091410

Pluye, P., Granikov, V., Bartlett, G., Grad, R. M., Tang, D. L., Johnson-Lafleur, J., … Doray, G. (2014). Development and content validation of the information assessment method for patients and consumers. *JMIR Research Protocols*, *3*(1), 1–15. https://doi.org/10.2196/resprot.2908

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what is being reported? Critique and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9

Purwaningsih, W., Sari, S. P., & Suryadi, A. (2020). The effect of stem-pjbl and discovery learning on improving students ' problem-solving skills of the impulse and momentum topic. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 465–476. https://doi.org/10.15294/jpii.v9i4.26432

Redaksi. (2024). (Siaran Pers) *Temuan riset mafindo: mayoritas warga belum bisa bedakan hoaks dan fakta* [(press release) mafindo research findings: the majority of citizens cannot distinguish between hoaxes and facts].

Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian (panduan peneliti, mahasiswa, dan psikometrian) (i) [quantitative analysis of research instruments (guide for researchers, students, and psychometricians) (I)]* (I). Yogyakarta: Parama Publishing.

Rodić, D. D. (2018). Best practices of assessment as a way to promote effective learning. *Journal of Baltic Science Education*, *17*(5), 748–750. https://doi.org/10.33225/jbse/18.17.748

Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement*, *40*(2), 163–184. https://doi.org/10.1111/j.1745-3984.2003.tb01102.x

Ropii, M., & Fahrurrozi, M. (2017). *Evaluasi hasil belajar* [evaluation of learning outcomes]. In *Yogyakarta: Pustaka Pelajar*.

Samsudin, H., Sadiman, & Pachrozi, I. (2019). *Kajian sosial dan pemerintahan berbasis geospasial bidang pendidikan (sistem informasi pendidikan berbasis geospasial) [*geospatial-based social and government studies in the field of education (geospatial-based education information Sys*)]* (S. Franeka, Ed.). Banyuasin:

Penerbit Bappeda Litbang.
Schwab, K. (2016). *The fourth industrial revolution* (Vol. 11). World Economic Forum.
Shaffer, J. F., Ferguson, J., & Denaro, K. (2019). Use of the test of scientific literacy skills reveals that fundamental literacy is an important contributor to scientific literacy. *CBE—Life Sciences Education*, *18*(3), ar31. https://doi.org/10.1187/cbe.18-12-0238
She, H. C., Stacey, K., & Schmidt, W. H. (2018). Science and mathematics literacy: pisa for better school education. *International Journal of Science and Mathematics Education*, *16*, 1–5. https://doi.org/10.1007/s10763-018-9911-1
Sidiq, M. N., & Permanasari, A. (2022). The analysis of STEM career interest of students aged 13-15 as an overview for the development of STEM career counseling. *ICONSEIR*. https://doi.org/10.4108/eai.21-12-2021.2317329
Stein, M. M., Smith, E. M., & Holmes, N. G. (2019). Confirming what we know: Understanding questionable research practices in intro physics labs. *2018 Physics Education Research Conference Proceedings*. American Association of Physics Teachers. https://doi.org/10.1119/perc.2018.pr.Stein
STEM, I. (2024). *Jumlah lulusan STEM indonesia jauh lebih rendah dibanding negara lain* [Indonesia's STEM graduates are much lower than other countries]. Retrieved July 12, 2025, from STEM Indonesia website: https://stem.or.id/2024/02/21/jumlah-lulusan-stem-indonesia-jauh-lebih-rendah-dibanding-negara-lain/
Sudirman, Rusilowati, A., & Susilaningsih, E. (2024). Development of multiplechoice test instruments to improve scientific literacy in madrasah aliyah (MA). *International Journal of Scientific Research and Management (IJSRM)*, *12*(06), 3465–3475. https://doi.org/10.18535/ijsrm/v12i06.el04
Suhaini, M., Ahmad, A., & Bohari, N. M. (2021). Assessments on vocational knowledge and skills: a content validity analysis. *European Journal of Educational Research*, *volume-10-*(volume-10-issue-3-july-2021), 1529–1540. https://doi.org/10.12973/eu-jer.10.3.1529
Suwarna, I. P., & Fatimah. (2018). *Implementasi penugasan digital untuk meningkatkan higher order thinking skills (hots) pada kemampuan siswa menengah atas dalam konsep hukum newton [implementation of digital assignments to improve higher order thinking skills (HOTs) Ability of Senior High*. *10*(2), 335–340. https://doi.org/http://dx.doi.org/10.15408/es.v10i2.10236
Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2
Teig, N. (2020). Scientific inquiry in TIMSS and PISA 2015. Inquiry as an instructional approach and the assessment of inquiry as an instructional outcome in science. *Nordic Studies in Science Education*, *16*(2), 235. https://doi.org/10.5617/nordina.8029
Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). Instructional development for training teachers of exceptional children: A sourcebook. In *Bloomington: Indiana University*. https://doi.org/10.1016/0022-4405(76)90066-2
Townley, A. (2018). Teaching and learning science in the 21st century: challenging critical assumptions in post-secondary science. *Education Sciences*, *8*(1), 12.

https://doi.org/10.3390/educsci8010012
Valladares, L. (2021). Scientific literacy and social transformation. *Science & Education*, *30*, 557–587. https://doi.org/10.1007/s11191-021-00205-2
Vázquez-López, V., & Huerta-Manzanilla, E. L. (2021). Factors related with underperformance in reading proficiency, the case of the programme for international student assessment 2018. *European Journal of Investigation in Health, Psychology and Education*, *11*(3), 813–828. https://doi.org/10.3390/ejihpe11030059
Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología*, *33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401
Walsh, C., Quinn, K. N., & Holmes, N. G. (2019). Assessment of critical thinking in physics labs: concurrent validity. *2018 Physics Education Research Conference Proceedings*. American Association of Physics Teachers. https://doi.org/10.1119/perc.2018.pr.Walsh
Yuberti, Y., Sairi, A. P., Nanto, D., & Sholeha, S. (2020). Physics Ludo integrated with scientific literacy as a Newton's law learning media. *Journal of Physics: Conference Series*, *1572*(1), 012051. https://doi.org/10.1088/1742-6596/1572/1/012051
Zhang, L., Liu, X., & Feng, H. (2023). Development and validation of an instrument for assessing scientific literacy from junior to senior high school. *Disciplinary and Interdisciplinary Science Education Research*, *5*(1), 21. https://doi.org/10.1186/s43031-023-00093-2
Zulfiani, Z., Permana Suwarna, I., Muin, A., Mulyati, T., & El Islami, R. A. Z. (2023). Developing the MathSci 21st app: Enhancing higher-order thinking skills assessment in mathematics and science education within an Islamic context. *International Journal of Advanced and Applied Sciences*, *10*(8), 19–31. https://doi.org/10.21833/ijaas.2023.08.003

- **APPENDIX**

Appendix 1. Example of scientific literacy questions
Wireless charging stimulus:

Wireless Charging

Sumber: https://productnation.co/id/15208/wireless-charger-terbaik-indonesia/

Wireless charging (pengisian daya secara nirkabel) didasarkan pada konsep medan gelombang elektromagnetik dan kuatnya kopling dari resonansi magnetik, yang menggunakan sifat induksi elektromagnetik yang ditimbulkan oleh kumparan pengirim dan penerima, sehingga dapat menghantarkan tegangan dan arus listrik. Di pasaran, wireless charging umumnya memiliki input tegangan antara 5-20 Volt dengan arus 3,25 A. Wireless charging yang optimal tercapai dengan dukungan daya besar dan efisiensi transfer energi yang tinggi, sehingga baterai dapat terisi lebih cepat dan minim kehilangan daya. Wireless charging aman digunakan karena tidak adanya kontak logam secara langsung. Terdapat dua jenis Wireless charging yang sering digunakan untuk mengisi daya handphone, yaitu Wireless Qi Standard dan Wireless Fast Charging. Keduanya mengisi daya tanpa kabel, tetapi dengan kecepatan yang berbeda. Perbedaan kecepatan ini salah satunya dipengaruhi oleh efisiensi transfer daya. Dilakukan sebuah pengujian untuk mengetahui perbedaan kecepatan dan efisiensi dua buah wireless, agar kita dapat memilih wireless charger yang tepat untuk digunakan setiap hari.

Prosedur A (15 Watt):
- *Pastikan handphone memiliki daya 0%*
- *Ukur presentase pengisian baterai tiap 10 menit*
- *Letakkan handphone pada ruangan bersuhu 30℃*
- *Ulangi sebanyak 3 kali*
- *Pasang termometer digital dekat ponsel untuk mencatat suhu*
- *Catat daya input menggunakan wattmeter*

Prosedur B (5 Watt):
- *Pastikan handphone memiliki daya 0%*
- *Ukur presentase pengisian baterai tiap 10 menit*
- *Letakkan handphone pada ruangan bersuhu 30℃*
- *Ulangi sebanyak 3 kali*
- *Pasang termometer digital dekat ponsel untuk mencatat suhu*
- *Catat daya input menggunakan wattmeter*

Reno melakukan eksperimen mengenai sistem wireless charging dengan spesifikasi sebagai berikut: Berdasarkan eksperimen yang dilakukan Reno, didapatkan hasil eksperimen sebagai berikut:

Tabel 1. *Spesifikasi rangkaian*

Jenis Rangkaian	***Panjang lilitan (m)***	***Jumlah lilitan***	***Kapasitor (Farad)***	***Luas Penampang (mm^2)***
Pemancar	0.3	1	40.8×10^{-9}	28.26
Penerima	0.3	1	34×10^{-9}	28.26

Tabel 2. *Hasil perbandingan seluruh pengujian*

1	***Daya Pemancar (W)***	***Penerima***					
		Beban Lampu 100 watt		***Beban Lampu 70 watt***		***Beban Lampu 25 watt***	
		P (W)	***Efisiensi (%)***	***P (W)***	***Efisiensi (%)***	***P (W)***	***Efisiensi (%)***
10	90.06	0.25	0.28	1.56	1.73	2.05	2.28

8	90.06	0.6	0.67	4.71	5.23	6.03	6.7
6	90.06	1.87	2.07	13.3	14.8	16.3	18.1
4	90.06	9.2	10.2	36.7	40.7	44.3	49.2
2	90.06	35.8	39.7	70.9	78.7	83.7	92.9

Example question on wireless charging (question code: L1.1):

Rere mengisi daya ponselnya menggunakan wireless charging x selama 45 menit hingga baterai penuh. Sementara itu, Mira hanya membutuhkan 30 menit untuk mengisi penuh baterainya menggunakan wireless chaging y. Jika persentase awal baterai dan spesifikasi ponsel mereka sama, mengapa pengisian daya pada ponsel Mira lebih cepat dibandingkan ponsel Rere?

Earthquake stimulus:

Gempa Bumi

Sumber: https://tirto.id/usai-gempa-dan-tsunami-di-palu-kenapa-komunikasi-ponsel-bermasalah-c3PY

Gempa bumi adalah getaran pada permukaan bumi akibat pelepasan energi secara tiba-tiba dari dalam bumi, umumnya dipicu oleh pergerakan lempeng tektonik. Intensitasnya bervariasi, mulai dari yang lemah dan hanya terdeteksi seismometer, hingga yang terasa kuat. Energi dilepaskan dari pusat gempa dan merambat ke segala arah dalam bentuk gelombang seismik. Dua jenis gelombang utama yang tercatat oleh seismograf adalah Gelombang P (primer) yang tercepat, dan Gelombang S (sekunder) yang lebih lambat. Selisih waktu kedatangan kedua gelombang ini dimanfaatkan untuk memperkirakan jarak pusat gempa, yang merupakan bagian dari metode ilmiah dalam seismologi.

Gempa bumi dapat menimbulkan kerusakan pada bangunan, jalan, lahan, dan jaringan listrik, serta memakan korban jiwa. Contohnya, gempa di NTB, Bali, dan sekitarnya menyebabkan pemadaman di beberapa pembangkit listrik seperti PLTU Jeranjang 25 MW, PLTU IPP LED 2x25 MW, PLTD MFO Cogindo 2x7 MW, PLTD Taman 0,65 MW, dan PLTD Paokmotong 5 MW. Dalam pemulihannya, pembangkit yang lebih stabil seperti PLTU harus diprioritaskan karena menjadi tulang punggung sistem, sementara PLTD berperan sebagai cadangan sementara.

Selain kerusakan infrastruktur, gempa bumi juga sering dikaitkan dengan gangguan sistem komunikasi dan navigasi, seperti yang terjadi pada gempa di Sulawesi Utara. Sebagian peneliti mengaitkan gangguan tersebut dengan aktivitas geomagnetik yang ditunjukkan oleh nilai Indeks-A yang tinggi (80) pada hari kejadian. Mereka berpendapat bahwa fluktuasi medan magnet bumi memicu pelepasan energi di zona subduksi, sehingga menyebabkan gempa. Namun, ahli geofisika membantah pandangan ini. Mereka menjelaskan bahwa Indeks-A yang tinggi mencerminkan aktivitas badai matahari, bukan proses tektonik. Gangguan komunikasi dan navigasi lebih mungkin disebabkan oleh dampak badai matahari terhadap ionosfer, sementara pemadaman listrik diakibatkan oleh kerusakan fisik pada pembangkit akibat guncangan gempa. Kedua peristiwa tersebut terjadi bersamaan secara kebetulan tanpa hubungan sebab-akibat. Nilai Indeks-A harian digunakan untuk menentukan kriteria aktivitas geomagnet harian. Berikut tabel kriteria aktivitas geomagnet berdasarkan rentang Indeks A.

Tabel 1. *Kriteria aktivitas geomagnet*

Rentang Indeks A	***Kriteria Aktivitas Geomagnet***
$0 \leq A < 4$	*Relatif tenang*
$4 \leq A < 7$	*Ada sedikit gangguan magnet*
$7 \leq A < 15$	*Ada gangguan magnet*
$15 \leq A < 29$	*Ada gangguan magnet aktif*
$29 \leq A < 50$	*Badai magnet kecil*
$50 \leq A < 101$	*Badai magnet besar*
$101 \leq A < 208$	*Badai magnet kuat*
$A \geq 208$	*Badai magnet sangat besar*

Example question on earthquakes (question code: G2.2):

Badan Meteorologi Klimatologi dan Geofisika (BMKG) mengamati waktu kedatangan gelombang P dan S di Sulawesi Tengah. Pada jarak 300 km dari pusat gempa, gelombang P tercatat setelah 50 detik, sedangkan gelombang S setelah 71,4 detik. Berdasarkan data tersebut, apa yang Anda dapat simpulkan mengenai penerapan metode ilmiah dalam mempelajari gempa bumi?

a) Gelombang P yang datang lebih dulu menunjukkan bahwa intensitas gempa diukur berdasarkan gelombang pertama yang tercatat.
b) Perbedaan waktu kedatangan gelombang P dan S pada jarak yang sama tidak berpengaruh dan dapat diabaikan dalam penyelidikan.
c) Waktu kedatangan gelombang S lebih penting karena dapat memperkirakan kekuatan gempa, sehingga proses ilmiah hanya fokus pada gelombang S.
d) *Data selisih waktu kedatangan gelombang P dan S dapat digunakan untuk memperkirakan jarak pusat gempa, yang merupakan bagian dari metode ilmiah dalam seismologi.*

Maglev train stimulus:

Kereta Maglev

Sumber: https://redigest.web.id/2024/03/proyek-maglev-chuo-shinkansen-diprediksi-ngaret/#google_vignette

Kereta Maglev (Magnetic Levitation) merupakan contoh penerapan konsep fisika dalam transportasi modern. Teknologi ini menggunakan gaya magnet untuk mengangkat dan menggerakkan kereta tanpa roda, sehingga bisa melaju hingga 500 km/jam. Karena tidak menyentuh rel, gesekan menjadi sangat kecil, perjalanan lebih halus dan tidak menghasilkan polusi udara. Namun, Maglev membutuhkan listrik dalam jumlah besar untuk menciptakan medan magnet yang kuat, sehingga biaya operasionalnya sangat tinggi. Selain itu, teknologi ini memerlukan jalur khusus yang harus dibangun dari awal, sehingga butuh biaya dan perencanaan yang lebih besar. Selain itu, ada potensi kebisingan yang lebih besar dan kebutuhan akan lintasan yang lurus. Sistem Maglev terdiri dari lima komponen utama, yaitu levitasi, sistem penggerak, pemandu, transfer daya input, dan kontrol.

Terdapat tiga sistem utama levitasi magnetik yang digunakan di dunia:

1. **Electro-Dynamic Suspension (EDS):** *Sistem ini memanfaatkan magnet superkonduktor yang sangat kuat dan menggunakan pemandu gaya tolak magnet. Karena stabil pada kecepatan tinggi sistem ini cocok untuk kereta jarak jauh. Namun, EDS memerlukan sistem pendingin kriogenik agar magnet tetap dalam suhu sangat rendah serta roda bantu untuk memulai gerakan hingga levitasi bisa terjadi.*
2. **Permanent Magnet Electro-Dynamic Suspension (PM-EDS):** *Sistem ini menggunakan magnet permanen yang bekerja pada suhu ruangan sehingga lebih hemat energi dan tidak butuh pendinginan khusus serta menggunakan pemandu yang sama seperti EDS. Tetapi, karena medan magnet yang dihasilkan tidak sepenuhnya stabil bisa muncul getaran atau osilasi kecil yang memengaruhi kenyamanan perjalanan.*
3. **Electromagnetic Suspension (EMS):** *Teknologi ini menggunakan elektromagnet biasa memungkinkan kereta melayang bahkan saat dalam keadaan diam dan menggunakan pemandu gaya tarik menarik magnet. Karena medan magnetnya lebih lemah, penumpang merasa lebih nyaman. Sistem EMS dengan sirkuit levitasi dan pemandu terpisah cocok untuk operasi kecepatan tinggi karena tidak adanya interfensi antara kedua sirkuit. Namun, menambah biaya desain.*

Ketika melintas, kereta ini juga mendemonstrasikan perubahan frekuensi suara yang terdengar akibat perbedaan posisi pendengar terhadap sumber bunyi yang bergerak cepat. Di negara dengan kondisi geografis seperti Indonesia, penerapan teknologi memerlukan kajian mendalam mengenai stabilitas sistem & efisiensi energinya.

Maglev train example question (question code: M3.3):

Suatu tim peneliti ingin mengembangkan sistem transportasi modern menggunakan teknologi kereta Maglev berkecepatan tinggi. Berikut tabel analisis komparatif sistem penggerak motor linier dinilai skala 1-5 untuk karakteristik tertentu. Nilai 5 menunjukkan respon terbaik, sedangkan 1 menunjukkan respon terburuk.

Tabel 1. *Analisis komparatif sistem penggerak motor linier*

Karakteristik	Sistem Penggerak			
	Brushless Direct Current Motor (BLDC)	Linear Induction Motor (LIM)	Linear Synchronous Motor (LSM)	Linear Switched Reluctance Motor (LSRM)
Rentang kecepatan	4	4	5	5
Biaya	4	3	4	5
Masa pakai	4	5	4	4.5

Buatlah rancangan kereta Maglev dengan menentukan jenis levitasi, sistem penggerak, dan pemandu kereta Maglev berkecepatan tinggi! Sertakan penjelasan atau alasan pemilihan setiap sistem berdasarkan informasi dari pernyataan diatas & tabel 1!

Appendix 2. Sample assessment rubric

Kode Soal	Kriteria Penilaian	Skor	Skor Maksimal
L1.1	*Memberikan alasan: lebih cepat karena charger yang digunakan Mira memiliki daya yang lebih besar dibandingkan dengan charger yang digunakan Rere*	1	2
	Memberikan alasan: lebih cepat karena charger yang digunakan Mira memiliki efisiensi yang lebih tinggi dibandingkan dengan charger yang digunakan Rere	1	
G2.2	*Memilih jawaban yang sesuai dengan kunci jawaban*	1	1
M3.3	*Menuliskan jenis sistem levitasi yaitu EMS dengan sirkuit levitasi dan panduan yang terpisah*	1	6
	Memberikan alasan dengan benar: tidak ada interfensi antara sirkuit levitasi dan pemandu	1	
	Menuliskan jenis sistem penggerak yaitu LSM dan LSRM	1	
	Meberikan alasan dengan benar: berdasarkan nilai skala rentang kecepatan sistem LSM dan LSRM menghasilkan nilai skala yang paling besar yaitu 5	1	
	Menuliskan jenis pemandu yaitu pemandu gaya tarik-menarik magnet	1	
	Meberikan alasan dengan benar: sistem EMS menggunakan pemandu gaya tarik-menarik magnet	1	