e-ISSN 3090-5354
Vol.2, No.2, August 2026

PISA 2025-Based Digital Physics Literacy Assessment: A Development Study on Wave Material in Lamongan

Sri Dewi[1*], Muhammad Satriawan[1], Titin Sunarti[1], Iqbal Ainur Rizki[2]

[1]State University of Surabaya, Surabaya, Indonesia
[2]Victoria University of Wellington, Wellington, New Zealand

Check for updates OPEN ACCESS CC BY

DOI : https://doi.org/10.26740/jdpe.2.2.55537

Sections Info

Article history:
Submitted: April 02, 2026
Final Revised: August 15, 2026
Accepted: August 15, 2026
Published: August 27, 2026

Keywords:
Assessment Digital;
HOTS;
Physics Literacy;
PISA 2025;
Wave Material.

ABSTRACT

Objective: *This study aimed to develop and validate a digital assessment instrument for physics literacy, framed around the three PISA 2025 science competencies, using wave material for grade XI students at a senior high school in Lamongan.* ***Method:*** *The study used a Research and Development (R&D) approach following the ADDIE model, limited to the Development stage, comprising expert validation and a limited trial involving 35 grade XI Phase F students. The instrument consisted of 10 higher-order-thinking essay items (Bloom's cognitive levels C4–C6), with a maximum score of 32 points, contextualized around the Sumenep Earthquake, East Java (M6.0, 30 September 2025).* ***Results:*** *The instrument was very valid (CVI = 0.89) and reliable (Cronbach's α = 0.84). Students' physics literacy profile was distributed as 45.7% Developing, 34.3% Capable, and 20.0% Proficient, with a class average of 76.0% (Capable category). Achievement per PISA 2025 competency was K-1, Explaining Scientific Phenomena, 80.3%; K-2, Evaluating Investigation Design and Interpreting Data, 74.9%; and K-3, Evaluating Scientific Information and Decision-Making, 73.4%.* ***Novelty:*** *This is the first PISA 2025-based digital assessment instrument for physics literacy contextualized on an authentic local seismic phenomenon in East Java, validated by a multi-expert panel comprising physics-education academics and a practicing teacher.*

INTRODUCTION

The ongoing digital transformation of education, along with the growing demand for twenty-first-century competencies such as critical thinking, evidence-based reasoning, and information evaluation, requires assessment practices that move beyond routine recall to measure students' capacity to reason with authentic, technology-mediated data (Sung et al., 2022). Within this shift, scientific literacy, and physics literacy as one of its central domains, has become a key indicator of whether learners can apply disciplinary knowledge to real-world problems rather than merely reproduce formulas. Physics literacy is defined as students' ability to understand, interpret, and apply physics concepts to scientifically explain natural phenomena, evaluate investigation designs, and use scientific information in real-life decision-making (Holbrook & Rannikmäe, 2009), consistent with the broader principle that contextualizing science education strengthens meaningful learning (Holbrook & Rannikmäe, 2020). As a key domain of scientific literacy, physics literacy is a primary focus of the Program for International Student Assessment (PISA), which measures students' competence in applying physics knowledge to real-life situations (OECD, 2024), an area that has seen rapidly growing global research attention over the past decade, as confirmed by systematic literature analyses following PRISMA guidelines (Alhusni et al., 2024; Suryawati et al., 2023).

The 2022 PISA results ranked Indonesia 67th in science, with an average score of 383, well below the OECD average of 485 (OECD, 2023). This gap reflects several interacting causes, including limited instructional time devoted to inquiry, uneven teacher familiarity with literacy-based pedagogy, and, of direct relevance to the present study, an assessment culture that rarely requires students to explain phenomena, evaluate investigation designs, or judge the credibility of scientific claims — the very competencies that scientific literacy demands (Astutik et al., 2022; Kemendikbudristek, 2023). In other words, weak literacy-oriented assessment is not the only explanation for Indonesia's PISA performance, but it is one of the more tractable ones, because it is a variable that classroom-level instrument development can directly address. A needs analysis at a senior high school in Lamongan confirmed that the available assessment instruments did not comprehensively measure physics literacy competencies according to the PISA framework, with questions still dominated by low cognitive levels (Astutik et al., 2022).

The PISA 2025 framework defines three core science competencies: K-1 (explaining phenomena scientifically), K-2 (developing and evaluating investigation designs and critically interpreting data), and K-3 (evaluating and reflecting on scientific information for decision-making) (OECD, 2024). It is important to clarify how these literacy competencies relate to higher-order thinking skills (HOTS), since the two are sometimes conflated in the physics-education literature. HOTS, defined through the Analyzing, Evaluating, and Creating levels of the Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), is not itself the construct this study measures; rather, it is the cognitive demand through which physics literacy is expressed. Explaining a phenomenon scientifically (K-1) requires analysis of causal mechanisms; evaluating an investigation design (K-2) requires evaluative judgment of methodological choices; and judging the credibility of scientific information (K-3) requires the synthesis and creation of a reasoned position. Because none of these performances can be produced through recall or routine application alone, PISA 2025-aligned literacy items are necessarily written at the C4–C6 levels (Widana, 2017). This study therefore treats the HOTS items as the instrument's cognitive vehicle, and physics literacy operationalized as K-1, K-2, and K-3 as the construct under measurement, consistent with prior work demonstrating the feasibility of developing valid HOTS-based physics instruments (Arafah et al., 2021; Bhakti et al., 2023; Sugiarti et al., 2017) and with evidence that well-designed HOTS instruments can meaningfully strengthen students' scientific and digital literacy (Astalini et al., 2022).

Over the past five years, several Indonesian and regional studies have pursued digital or multimedia-based physics instruments and learning media, but a closer look at this body of work reveals a persistent gap that motivates the present study. Digital books and 3D animation embedded in online problem-based learning have been shown to strengthen pre-service teachers' scientific literacy (Sunarti et al., 2024); interactive multimedia on sound and light waves has improved students' concept understanding and literacy skills (Asrizal et al., 2025); STEM-based e-learning has produced measurable communication and collaboration gains on wave topics (Wiyono et al., 2022); and website- or STEAM-based platforms have been developed to support physics learning more generally (Sina et al., 2023; Mahendra et al., 2025). Assessment-focused instrument development has likewise progressed, including PISA-model instruments for problem-solving and scientific literacy at the junior-secondary level (Rosana et al., 2020), general science-literacy profiling instruments (Chasanah et al., 2022), and reviews contrasting the PISA 2018 and PISA 2025 science frameworks (Deta et al., 2024). What remains absent from this literature, however, is a digital instrument that (a) operationalizes the specific three-competency structure of PISA 2025 rather than scientific literacy in general, (b) is anchored to an authentic, dated local phenomenon rather than a generic or hypothetical scenario, and (c) is validated by a panel that combines academic and school-practitioner perspectives. This combination constitutes the gap the present study addresses, and is the basis of its claimed novelty.

Digital assessment offers advantages in the form of integrating multimedia stimuli such as seismograms and data graphs, as well as more systematic data collection (Lumembang & Bandari, 2024), in line with the transformation of the Independent Learning Curriculum (Kemendikbudristek, 2023). Several recent studies have shown the effectiveness of digital assessment in increasing the authenticity of physics learning-outcome measurements (Mahendra et al., 2025; Sina et al., 2023), consistent with broader evidence that technology-supported learning strengthens higher-order thinking (Sung et al., 2022) and that problem-based approaches improve physics problem-solving abilities (Putri et al., 2024).

Wave material was chosen as the assessment context because it encompasses physics concepts wave propagation, energy relationships, resonance, and interference that lend themselves naturally to literacy-based tasks requiring analysis, evaluation, and creation, rather than to computation alone (Fadilah et al., 2020). This choice is consistent with prior work developing valid physics assessment instruments (Simamora et al., 2024) and demonstrating measurable literacy and skill gains specifically on wave material (Wiyono et al., 2022; Asrizal et al., 2025), and with evidence that localized, real-world contexts strengthen literacy-oriented achievement in physics learning (Martawijaya et al., 2023). To strengthen contextual relevance, the Sumenep Earthquake phenomenon of 30 September 2025 (M6.0) was used as an authentic stimulus, given that Indonesia records more than 6,000 earthquakes per year (BMKG, 2025), which underscores the importance of embedding disaster-mitigation understanding within the school curriculum (Putra et al., 2021; Rofiah et al., 2021). One senior high school in Lamongan was chosen as the research site because it has adequate digital infrastructure, while its available assessment instruments do not yet measure the three PISA 2025 competencies in an integrated manner (Chasanah et al., 2022), a gap consistent with similar efforts elsewhere to develop PISA-model instruments for measuring problem-solving skills and scientific literacy (Rosana et al., 2020).

Based on this background, this study aims to develop a digital assessment instrument for physics literacy based on PISA 2025 on wave material, with a maximum score of 32 points, while also describing the instrument's content validity, drawn from the judgment of two physics-education lecturers and one high-school physics teacher, together with its reliability, and describing the physics literacy profile of grade XI students per PISA 2025 competency, so that these three aims jointly produce not only a validated measurement tool but also an empirical baseline of students' strengths and weaknesses across the K-1, K-2, and K-3 competencies; the practical implication is that teachers and curriculum developers at the research site, and potentially at comparable schools, gain both a ready-to-use, locally contextualized assessment instrument and evidence-based guidance on which competency namely critical evaluation and argument construction under K-3 most urgently needs to be strengthened through subsequent instructional design.

RESEARCH METHOD

This study used the Research and Development (R&D) method with the ADDIE model developed by Branch (2009), consistent with the broader R&D tradition in educational research (Gall et al., 2007). ADDIE was chosen over alternative R&D designs, such as the 4D model, because its five iterative, criterion-referenced stages allow validity and reliability evidence to be generated as early as the Development stage, before a resource-intensive, full-scale implementation. This sequencing matched the present study's aim, which was to establish the instrument's initial feasibility rather than test its classroom-wide effectiveness. The ADDIE (Analysis, Design, Development, Implementation, and Evaluation) of which this study covers the first three (Analysis, Design, and Development); the Implementation and Evaluation stages were intentionally deferred rather than omitted. Administering an unvalidated instrument across multiple classes or schools would risk producing an unreliable picture of students' physics literacy, so establishing content validity and internal-consistency reliability was treated as a necessary prerequisite before broader classroom implementation could be justified. These two remaining stages are planned for a subsequent study with a broader subject scope, so that the product's effectiveness can be tested comprehensively. Figure 1 presents the research flow used in this study.

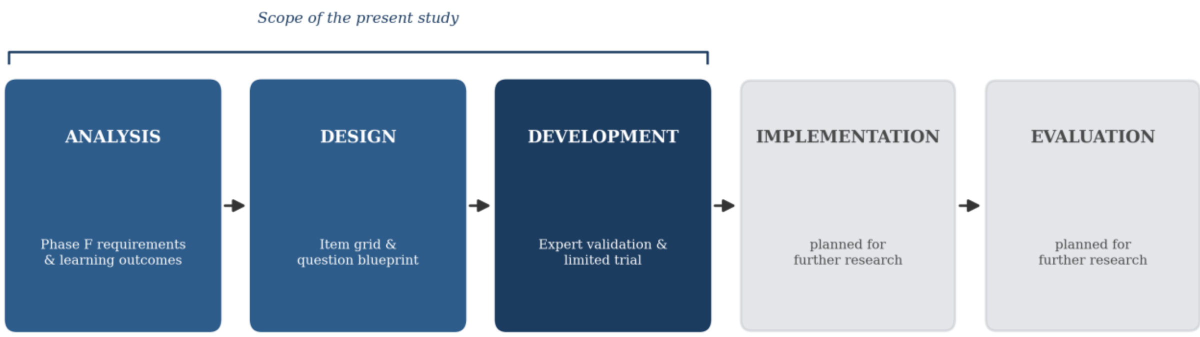

Figure 1. ADDIE model development flowchart (Branch, 2009)

Development Procedures

The three ADDIE stages implemented in this study proceeded as follows.

1. Analysis stage: needs analysis through teacher interviews and classroom observations, together with an analysis of the alignment between Phase F learning outcomes and the PISA 2025 competencies (OECD, 2024).
2. Design stage: preparation of the item characteristics and grid for 10 open-ended items distributed across K-1 (items S1, S2, S5), K-2 (items S3, S4, S6, S7), and K-3 (items S8, S9, S10), written at cognitive levels C4–C6, with a maximum score of 32 points, using the Sumenep M6.0 earthquake as the contextual phenomenon.
3. Development stage: drafting of the test items, analytical scoring rubrics, and answer keys, followed by digitization of the instrument via Google Forms (Lumembang & Bandari, 2024), and then expert validation and a limited trial.

Subjects and Place of Research

The research was conducted in April 2026 at a State Senior High School (SMAN) in Lamongan Regency. The subjects in the limited-trial phase were 35 grade XI Phase F students who had completed the wave-material unit under the Independent Curriculum. Subjects were selected purposively, based on the physics teacher's recommendation, to ensure that the material students had studied matched the context of the instrument being developed (Sugiyono, 2019).

Validation Instrument

Three experts carried out the validation: two physics-education lecturers and one high-school physics teacher. Each validator assessed all 10 items on three aspects using a 1–4 scale: (1) Content (K) the alignment of items with physics-literacy indicators and PISA 2025 competencies; (2) Construction (Kn) the clarity of item construction, alignment with cognitive levels C4/C5/C6, and consistency of the scoring rubric; and (3) Language (B) clarity, accuracy, and appropriateness of the language used.

Data Analysis Techniques

Content validity was analyzed using the Content Validity Ratio (CVR) and Content Validity Index (CVI). The CVR for each item was calculated using the Lawshe (1975) equation:

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}} \qquad (1)$$

Where n is the number of validators who assigned a score ≥ 3 (indicating the item is valid), and N is the total number of validators. The CVI was then calculated as the mean CVR across all items:

$$CVI = \frac{\Sigma CVR}{k} \qquad (2)$$

Where ΣCVR is the sum of the CVR values for all items, and k is the number of items. Item-level CVR critical values were determined based on the number of validators (Ayre & Scally, 2014), an approach comparable to other expert-judgment content-validity methods such as Aiken's V (Kania et al., 2024). An instrument is considered highly valid if it achieves a CVI > 0.80 (Chasanah et al., 2022). Table 1 presents the instrument's eligibility criteria based on CVI.

Table 1.CVI eligibility criteria

CVI Value	Category	Description
> 0.80	Very Valid	Suitable for use without major revisions
0.60 – 0.80	Valid	Suitable for use with minor revisions
< 0.60	Not Valid	Needs a complete revision

Instrument reliability was analyzed using Cronbach's alpha with SPSS version 26. The instrument was considered reliable if its reliability coefficient (α) was ≥ 0.70, consistent with commonly applied cut-off criteria for internal consistency in educational research (Zakariya, 2022), complementing item-level approaches such as Rasch-model analysis used for other physics-concept instruments (Kurniawan et al., 2024). After confirming the instrument's validity and reliability, the limited-trial results were analyzed descriptively to profile students' physics literacy. Students' percentage scores were categorized into four levels, as shown in Table 2.

Table 2. Student literacy profile category

Percentage Score (%)	Category
> 80	Proficient
61 – 80	Capable
41 – 60	Developing
≤ 40	Beginning

RESULTS AND DISCUSSION

Results

Development of the Assessment Instrument

The digital physics-literacy assessment instrument consists of two authentic reading texts based on the M6.0 Sumenep Earthquake (30 September 2025), four visual stimuli (a seismogram, a diagram of seismic wave types, an earth cross-section, and a building-resonance graph), and 10 open-ended items with a maximum score of 32 points (BMKG, 2025). The instrument measures the three PISA 2025 competencies (K-1, K-2, K-3) through three Bloom's cognitive levels (C4, C5, C6). Based on the needs analysis, Table 3 presents the item characteristics.

Table 3. Characteristics of the digital physics literacy assessment based on PISA 2025

Characteristic	Description
Item type	Open-ended items based on authentic text and multi-representational visual stimuli
Stimulus form	Multi-representation: BMKG news text, seismogram, seismic-wave-type diagram, earth cross-section, and building-resonance graph

Characteristic	Description
Context	Local, authentic context: the Sumenep Earthquake, East Java (M6.0, 30 September 2025), a real seismic phenomenon relevant to East Javanese learners
Physics content	Mechanical waves: P/S/L wave propagation, wave energy ($E \propto A^2$), epicenter distance, building resonance, and properties of the propagation medium
PISA 2025 competencies	K-1: Explaining Scientific Phenomena (3 items); K-2: Evaluating Investigation Design and Data Interpretation (4 items); K-3: Evaluating Scientific Information and Decision-Making (3 items)
Bloom's cognitive levels	C4 Analyzing: 6 items; C5 Evaluating: 2 items; C6 Creating: 2 items — all items fall within the HOTS range (C4–C6)
Number of items	10 open-ended items
Scoring system	Analytical rubric: 0–3 for standard items, 0–4 for complex items; maximum total score: 32
Delivery platform	Google Forms (digital/online); accessed via the school computer laboratory; 90 minutes to complete
Target respondents	Grade XI Phase F students who had completed the wave-material unit (Independent Curriculum)

Table 4 presents the item grid, combining each item's physics content, stimulus, literacy indicator, cognitive demand, Bloom level, and maximum score.

Table 4. Item Grid of the PISA 2025-Based Digital Physics Literacy Assessment

Item	Comp.	Physics content	Physics literacy indicator	Bloom	Score
S1	K-1	P, S, L seismic wave propagation	Sequences P, S, and L wave arrivals based on propagation speed in different media	C4	3
S2	K-1	Wave energy ($E \propto A^2$)	Calculates and proves the energy ratio of two waves from amplitude data	C4	3
S3	K-2	Epicenter distance	Uses $\Delta t = ts - tp$ and propagation speed to determine epicenter distance from BMKG data	C4	3
S4	K-2	Building resonance; natural period T0 vs. floor number N	Constructs a T0–N chart and identifies floor ranges most at risk of resonance	C4	3
S5	K-1	Properties of transverse-wave propagation in a liquid medium	Explains why liquid media, lacking shear-stress support, do not transmit S waves	C4	3
S6	K-2	Building resonance; simple pendulum as a model	Designs a systematic, scientifically valid investigation procedure (variables, tools, steps) for building resonance	C6	4
S7	K-2	Weaknesses of an experimental design	Identifies at least two methodological weaknesses and proposes physics-based improvements	C5	3
S8	K-3	Tsunami mechanism; thrust fault; hypocenter	Evaluates the validity of viral claims about the earthquake using seismological evidence	C5	3
S9	K-3	Predictability of aftershocks;	Explains the correlation–causality distinction and the limits of deterministic seismological prediction	C5	3

Item	Comp.	Physics content	Physics literacy indicator	Bloom	Score
		correlation–causality logic			
S10	K-3	Earthquake-resistant construction; wave–resonance integration	Integrates seismic data and resonance principles into evidence-based construction recommendations	C6	4

Expert Validation Results

Three validators assessed the instrument on the Content (K), Construction (Kn), and Language (B) aspects using a 1–4 scale. Table 5 presents the validation results together with the modal score and CVR value for each item.

Table 5. Expert validation results

Item	V1: K/Kn/B	V2: K/Kn/B	V3: K/Kn/B	Mode	CVR
1	4/4/4	4/3/4	3/4/4	4	1.00
2	4/3/4	4/3/4	3/3/4	4	0.89
3	4/4/3	4/3/3	3/4/3	3	0.89
4	3/4/4	3/4/3	3/3/4	3	0.89
5	4/3/3	4/3/3	3/3/3	3	0.78
6	4/4/4	4/4/4	3/4/4	4	1.00
7	3/4/3	3/4/3	3/3/3	3	0.78
8	4/4/4	4/4/4	3/4/4	4	1.00
9	3/3/4	3/3/3	3/3/3	3	0.78
10	4/4/4	4/4/4	3/4/4	4	1.00

All 10 items achieved a CVR ≥ 0.78, with seven items reaching CVR ≥ 0.89 (Very Valid). The overall instrument CVI was 0.89, in the Very Valid category, as it exceeded the 0.80 threshold (Chasanah et al., 2022). Items 5, 7, and 9 obtained CVR = 0.78 (Valid), each with a modal score of 3 across validators (Koza et al., 2024). Cronbach's alpha reliability was $\alpha = 0.84$, in the high-reliability category ($\alpha \geq 0.70$) (Ulu & Yerdelen-Damar, 2024).

Student Physics Literacy Profile

Table 6 presents the assessment scores of the 35 grade XI Phase F students, out of a maximum score of 32. No students fell into the Beginning category (≤ 40%), indicating that all students possessed a basic foundation of physics literacy (Holbrook & Rannikmae, 2009). The class average of 76.0% fell in the Capable category.

Table 6. Assessment scores of grade XI phase F students

No	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total	%	Category
1	3	3	2	3	3	3	3	3	2	3	28	87.5%	Proficient
2	3	2	2	2	2	3	3	3	2	3	25	78.1%	Capable
3	3	3	2	3	2	3	2	3	2	3	26	81.3%	Proficient
4	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
5	3	2	2	2	2	2	2	3	2	2	22	68.8%	Capable
6	3	3	2	3	3	3	3	3	3	3	29	90.6%	Proficient
7	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
8	3	2	2	3	2	3	2	3	2	3	25	78.1%	Capable

No	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total	%	Category
9	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
10	3	3	2	3	3	3	3	3	2	3	28	87.5%	Proficient
11	2	2	2	2	2	3	2	3	2	2	22	68.8%	Capable
12	3	3	2	3	2	3	2	3	2	2	25	78.1%	Capable
13	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
14	3	2	2	3	2	3	3	3	2	2	25	78.1%	Capable
15	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
16	3	3	3	3	3	3	2	3	2	3	28	87.5%	Proficient
17	2	2	2	2	2	3	2	3	2	2	22	68.8%	Capable
18	3	2	2	3	2	3	2	3	2	2	24	75.0%	Capable
19	2	2	2	2	2	2	2	3	2	2	21	65.6%	Capable
20	3	3	2	3	2	3	3	3	2	2	26	81.3%	Proficient
21	3	3	3	3	3	3	3	3	3	4	31	96.9%	Proficient
22	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
23	3	2	2	3	2	3	2	3	2	2	24	75.0%	Capable
24	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
25	3	3	3	3	3	4	3	3	3	4	32	100.0%	Proficient
26	2	2	2	2	2	3	2	3	2	2	22	68.8%	Capable
27	3	2	2	3	2	3	3	3	2	3	26	81.3%	Proficient
28	3	3	2	3	2	3	2	3	2	2	25	78.1%	Capable
29	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
30	3	3	3	3	3	3	3	3	3	4	31	96.9%	Proficient
31	2	2	2	2	2	2	2	3	2	2	21	65.6%	Capable
32	3	2	2	3	2	3	2	3	2	2	24	75.0%	Capable
33	2	2	2	2	2	2	2	2	2	2	20	62.5%	Capable
34	3	3	2	3	3	3	3	3	2	2	27	84.4%	Proficient
35	3	3	3	3	3	4	3	3	3	4	32	100.0%	Proficient

Distribution of Physics Literacy Categories

Table 7 summarizes the distribution of students' physics-literacy categories based on Table 6. The class average of 76.0% (Capable) is higher than results reported in previous Indonesian physics-literacy studies, which generally found most students at the Developing level (Deta et al., 2024). This is likely related to the strong personal relevance of the East Java seismic-phenomenon context for these students (Fadilah et al., 2020).

Table 7. Distribution of students' physics literacy categories

Category	Range (%)	n	%
Proficient	> 80	7	20.0%
Capable	61 – 80	12	34.3%
Developing	41 – 60	16	45.7%
Beginning	≤ 40	0	0%
Total		35	100%

Achievement of PISA 2025 Competencies

Table 8 presents achievement per PISA 2025 competency. Students performed best on K-1 (Explaining Scientific Phenomena: 80.3%), followed by K-2 (Investigation Design and

Data Interpretation: 74.9%) and K-3 (Evaluating Scientific Information and Decision-Making: 73.4%). This gradient aligns with scientific-literacy research in Indonesia, which has identified critical-evaluation tasks as the most challenging for students (Deta et al., 2024; Astutik et al., 2022).

Table 8. Average achievement per PISA 2025 competency

Competency	Items (max. score)	Mean (%)	Category
K-1: Explaining Scientific Phenomena	S1(3)+S2(3)+S5(3) = 9	80.3%	Capable
K-2: Investigation Design & Data Interpretation	S3(3)+S4(3)+S6(4)+S7(3) = 13	74.9%	Capable
K-3: Information Evaluation & Decision-Making	S8(3)+S9(3)+S10(4) = 10	73.4%	Developing

Discussion

Instrument validity and reliability. The validation results confirmed that the PISA 2025-based digital physics-literacy assessment on wave material was highly valid (CVI = 0.89) and reliable (α = 0.84), meeting the standards required for measuring physics literacy at the high-school level (Chasanah et al., 2022). The validators' modal scores, ranging from 3 to 4 across all items, reflected strong consensus. A validation panel of two physics-education lecturers and one high-school physics teacher strengthened the evaluation's credibility by integrating academic and practical perspectives (Astutik et al., 2022). Items 5, 7, and 9 obtained CVR = 0.78, mainly because the school-practitioner validator flagged language clarity, which illustrates why involving practitioners in HOTS-based instrument validation is important for balancing academic rigor with student readability (Koza et al., 2024).

Relationship between HOTS and physics literacy. The results reinforce the conceptual distinction drawn in the Introduction: HOTS is the cognitive mechanism, not the measured outcome. Item S6 (K-2, C6, designing an investigation) and item S10 (K-3, C6, generating evidence-based recommendations) required the highest cognitive level in the instrument and also carried the highest maximum scores (4 points each); students' relatively lower performance on the K-3 items generally (73.4%, the lowest of the three competencies) suggests that the Creating and Evaluating demands of HOTS are precisely where physics-literacy performance breaks down, not merely where item difficulty happens to be highest. This indicates that HOTS and physics literacy are not two separate constructs running in parallel but are functionally coupled: a student cannot demonstrate K-2 or K-3 literacy without also demonstrating the corresponding Bloom-level reasoning, and instrument developers should therefore treat HOTS items as a necessary but not sufficient condition for literacy-oriented assessment; item wording must still explicitly require explanation, evaluation, or the use of authentic evidence, or the item will measure analytical skill without measuring literacy.

The instrument directly addresses a gap identified during the analysis phase, namely the absence of a PISA 2025-based instrument that measures K-1 through K-3 in an integrated way at this Lamongan senior high school (Chasanah et al., 2022), and more broadly the gap identified in the Introduction — between the growing body of digital physics-learning media developed over the past five years (Asrizal et al., 2025; Mahendra et al., 2025; Sunarti et al., 2024; Wiyono et al., 2022) and the comparatively thin evidence base for digital instruments anchored specifically to PISA 2025's three-competency structure. Using the Sumenep Earthquake as the stimulus base generated authentic student responses, consistent with Fadilah et al. (2020), who found that contextualizing local disaster phenomena increases students' personal relevance to the assessment, and with recent findings on Indonesian pre-service physics teachers' seismic scientific literacy (Deta et al., 2025). The integration of BMKG news text, seismograms, and resonance graphs as stimuli reflects PISA 2025's emphasis on multi-representational literacy (OECD, 2024), while delivery through Google Forms provided practical advantages for visual-stimulus integration and data-collection efficiency, consistent with the growing use of digital, multimedia-assisted platforms in physics instruction (Prahani et al., 2022; Lumembang & Bandari, 2024).

Sources of the achievement gap. The class average of 76.0% (Capable) and the absence of students in the Beginning category indicate an adequate foundation of physics literacy. However, the achievement gradient from K-1 (80.3%) to K-2 (74.9%) to K-3 (73.4%) reveals a persistent gap in critical-evaluation competency. On item S8 (evaluating the viral claim that all undersea earthquakes trigger tsunamis), 71% of students correctly identified the claim as invalid. However, only 43% could articulate the specific physical mechanism involved: thrust-fault rupture, vertical seabed displacement, and hypocenter depth. On item S9 (critiquing the claim that aftershocks occur in a predictable pattern), 66% of students identified the underlying correlation–causality fallacy, but only 31% could explain why seismology cannot deterministically predict earthquake timing. This pattern indicates that students can recognize a questionable claim more readily than they can construct the evidence-based argument needed to refute it, underscoring the need for instruction that explicitly trains argumentation grounded in physical evidence, drawing on established frameworks such as Toulmin's Argument Pattern (Erduran et al., 2004). This gap mirrors comparable findings on students' critical-thinking and problem-solving skills in physics (Abeden & Siew, 2022; Suprapto et al., 2024), on analytical-rubric-based approaches to fostering critical thinking (Suryanti & Nurhuda, 2021), and on scientific-argumentation skills more broadly (Al-Ajmi & Ambusaidi, 2019; Astutik et al., 2022).

Learning models are suited to closing this gap. Because the deficit lies specifically in evaluation and argument construction rather than basic conceptual understanding, instructional approaches that have students practice constructing and defending evidence-based claims are more likely to close it than approaches that add more content coverage. Ethno-STEM project-based learning has been shown to improve HOTS and reduce misconceptions on locally contextualized physics topics (Martawijaya et al., 2023); problem-based learning has produced consistent gains in physics problem-solving across a recent meta-analysis (Putri et al., 2024); and online problem-based learning integrated with digital books and 3D animation has strengthened pre-service teachers' scientific literacy specifically (Sunarti et al., 2024). Explicit argumentation instruction using Toulmin's framework, combined with analytical-rubric feedback (Suryanti & Nurhuda, 2021), would target the K-3 weaknesses identified here directly, since it requires students to state a claim, ground it in evidence, and justify the connection between the two — the exact sequence on which items S8 and S9 revealed the largest gaps. Embedding these approaches within the wave-material unit, using the same kind of authentic local stimuli validated in this instrument, would let instruction and assessment reinforce the same competencies rather than the two remaining misaligned.

Physics literacy in the digital-transformation era. Beyond its role in the school curriculum, physics literacy of the kind measured here reading a seismogram, judging whether a viral claim about a disaster is scientifically grounded, and distinguishing correlation from causation in a stream of unverified reports is directly relevant to the demands of a digital-transformation era, in which citizens routinely encounter dashboards, sensor data, and rapidly circulating claims mediated by digital platforms (Sung et al., 2022; Astalini et al., 2022). An assessment instrument that measures this capacity digitally, using authentic multi-representational stimuli, therefore models the same evaluative literacy that students will need outside the classroom, and supports the broader argument that scientific literacy should be understood as a form of digital-era civic competence rather than a purely academic outcome (Binkley et al., 2012). A limitation to acknowledge is that scoring open-ended items still requires manual correction, which can introduce rater subjectivity; AI-assisted, rubric-referenced scoring could reduce this limitation while improving assessment efficiency and consistency (Mahendra et al., 2025).

CONCLUSION

Fundamental Findings: This research successfully developed a digital assessment instrument for physics literacy based on PISA 2025 on wave material, through the ADDIE model up to the Development stage, comprising 10 open-ended items that measure the three PISA 2025 competencies (K-1, K-2, K-3) at three Bloom cognitive levels (C4, C5, C6) with a maximum score of 32 points. Validation by three experts produced CVI = 0.89 (very valid) and Cronbach's $\alpha = 0.84$ (high reliability). The physics-literacy profile of the 35 grade XI students who trialed the instrument was 45.7% Developing, 34.3% Capable, and 20.0% Proficient, with a class average of 76.0% (Capable category) and a competency gradient of K-1 = 80.3%, K-2 = 74.9%, and K-3 = 73.4%. **Implications:** The results imply that the gap between K-1 and K-3 indicates a specific need for instruction that explicitly trains students to evaluate scientific information, identify logical fallacies, and construct arguments grounded in physical evidence, rather than a general need for more content coverage. Contextualizing authentic local disaster phenomena as an assessment stimulus proved effective here and merits extension to other physics topics through argumentation-focused and problem-based learning approaches; the validated instrument itself also offers teachers and curriculum developers a ready-to-use, locally contextualized tool for diagnosing students' physics-literacy profiles. **Limitations:** This study is limited to the Analysis, Design, and Development stages of the ADDIE model, applied to a single class at one senior high school in Lamongan, so the results cannot yet be generalized more broadly; the essay-based scoring format is also inherently subject to rater subjectivity. **Future Research:** Further research is recommended to carry out the Implementation and Evaluation stages comprehensively, expanding the subject pool to several senior high schools across Lamongan and East Java to test the instrument's effectiveness at scale, and to pursue the development of an AI-assisted automated scoring system to address the rater-subjectivity limitation inherent to essay-based assessment.

AUTHOR CONTRIBUTIONS

Sri Dewi contributed to the validation process, methodology development, data analysis, sourcing references, data management, and drafting the manuscript; **Muhammad Satriawan & Titin Sunarti** contributed to the conceptual framework, research design, and methodology development; **Iqbal Ainur Rizki** contributed to the conceptual framework, research design, and drafting the manuscript. All listed authors have reviewed and approved the final version of this submission.

CONFLICT OF INTEREST STATEMENT

The authors confirm that there are no conflicts of interest, either financial or personal, that may have influenced the content or outcome of this study.

ETHICAL COMPLIANCE STATEMENT

This manuscript complies with research and publication ethics. The authors affirm that the work is original, conducted with academic integrity, and free from any unethical practices, including plagiarism.

STATEMENT ON THE USE OF AI OR DIGITAL TOOLS IN WRITING

The authors acknowledge the use of digital tools, including AI-based technologies, as support in the research and writing stages of this article. Specifically, Claude was employed to assist with data analysis, and ChatGPT assisted with the development of the assessment instrument. All outputs generated with digital assistance were critically evaluated and revised to ensure academic rigor and ethical standards were upheld. The final responsibility for the manuscript rests entirely with the authors.

REFERENCES

Abeden, N. A. M., & Siew, N. M. (2022). Assessing students' critical thinking and physics problem-solving skills in secondary schools. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *7*(6), e001584. https://doi.org/10.47405/mjssh.v7i6.1584

Al-Ajmi, B., & Ambusaidi, A. (2019). The level of scientific argumentation skills in chemistry. *Science Education International*, *33*(1), 66–74. https://doi.org/10.33828/sei.v33.i1.7

Alhusni, H. Z., Habibbulloh, M., Lestari, N. A., Realita, A., Jatmiko, B., & Deta, U. A. (2024). Scientific literacy in physics learning: A bibliometric analysis from 1977 until

2023 and its impact on quality education. *E3S Web of Conferences*, *513*, 04012. https://doi.org/10.1051/e3sconf/202451304012

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Arafah, K., Amin, B. D., Sari, S. S., & Hakim, A. (2021). The development of higher order thinking skills (HOTS) instrument assessment in physics study. *Journal of Physics: Conference Series*, *1899*(1), 012140. https://doi.org/10.1088/1742-6596/1899/1/012140

Asrizal, A., Annisa, N., Ahzari, S., & Helma, H. (2025). Interactive multimedia sound and light waves integrated STEM to develop concept understanding and literacy skills of students. *Journal of Turkish Science Education*, *22*(1), 18–32. https://doi.org/10.36681/tused.2025.002

Astalini, A., Darmaji, D., Kurniawan, D. A., & Minarsih, M. (2022). Identification of HOTS creative thinking, science process skills and digital literacy in physics subject. *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, *12*(1), 47–61. https://doi.org/10.26740/jpfa.v12n1.p47-61

Astutik, S., Susantini, E., Madlazim, M., Suprapto, N., & Prahani, B. K. (2022). Evaluation of physical science literacy assessment instrument based on PISA 2025: A secondary school needs analysis. *Journal of Physics: Conference Series*, *2309*, 012054. https://doi.org/10.1088/1742-6596/2309/1/012054

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in physics: A systematic literature review. *Journal of Physics: Conference Series*, *2596*(1), 012067. https://doi.org/10.1088/1742-6596/2596/1/012067

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5_2

BMKG. (2025). *Laporan gempa tektonik M6,0 Sumenep, Madura, Jawa Timur, 30 September 2025*. Badan Meteorologi, Klimatologi, dan Geofisika.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

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

Deta, U. A., A'yun, S. K., Laila, L., Prahani, B. K., & Suprapto, N. (2024). PISA science framework 2018 vs 2025 and its impact in physics education: Literature review.

Momentum: Physics Education Journal, *8*(1), 95–107. https://doi.org/10.21067/mpej.v8i1.9215

Deta, U. A., Sasmi, R. R., Arisanti, A., Laila, L., Hudha, M. N., Mubarok, H., Prahani, B. K., & Suprapto, N. (2025). Earthquake scientific literacy profile of Indonesian first year pre-service physics teacher. *Journal of Education and Learning (EduLearn)*, *19*(2), 890–899. https://doi.org/10.11591/edulearn.v19i2.21624

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915–933. https://doi.org/10.1002/sce.20012

Fadilah, M., Permanasari, A., Riandi, R., & Maryani, E. (2020). Analisis karakteristik kemampuan literasi sains konteks bencana gempa bumi. *Jurnal IPA & Pembelajaran IPA*, *4*(1), 103–119. https://doi.org/10.24815/jipi.v4i1.16651

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson Education.

Holbrook, J., & Rannikmäe, M. (2020). Contextualized science education for meaningful learning. *International Journal of Science Education*, *42*(1), 1–18. https://doi.org/10.1080/09500693.2019.1701866

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. *REID (Research and Evaluation in Education)*, *10*(1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Kemendikbudristek. (2023). *Siaran pers: Peringkat Indonesia pada PISA 2022 naik 5–6 posisi dibanding 2018*. https://www.kemdikbud.go.id/main/blog/2023/12/peringkat-indonesia-pada-pisa-2022-naik-56-posisi-dibanding-2018

Koza, Y., Harso, A., & Doa, H. (2024). Pengembangan instrumen soal HOTS pada materi fluida statis. *OPTIKA: Jurnal Pendidikan Fisika*, *8*(1), 69–78. https://doi.org/10.37478/optika.v8i1.3555

Kurniawan, A., Istiyono, E., & Daeng Naba, S. (2024). Item quality analysis of physics concept understanding test with Rasch model. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *9*(3), 474–486. https://doi.org/10.26737/jipf.v9i3.5692

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

Lumembang, M. M., & Bandari, T. (2024). Utilitas platform digital pada pembelajaran fisika kelas X MAN Pinrang. *Al-Irsyad Journal of Physics Education*, *3*(1), 28–40. https://doi.org/10.58917/ijpe.v3i1.104

Mahendra, A., Prahani, B. K., & Kurtuş, M. A. (2025). Development of interactive physics learning website based on STEAM approach. *Journal of Digitalization in Physics Education*, *1*(1), 000003. https://doi.org/10.26740/jdpe.v1i1.38994

Martawijaya, M. A., Rahmadhanningsih, S., Swandi, A., Hasyim, M., & Sujiono, E. H. (2023). The effect of applying the ethno-STEM-project-based learning model on students' higher-order thinking skill and misconception of physics topics related to Lake Tempe, Indonesia. *Jurnal Pendidikan IPA Indonesia*, *12*(1), 1–13. https://doi.org/10.15294/jpii.v12i1.38703

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2024). *PISA 2025 science framework*. OECD Publishing. https://pisaframework.oecd.org/science-2025/

Prahani, B. K., Rizki, I. A., Nisa', K., Citra, N. F., Alhusni, H. Z., & Wibowo, F. C. (2022). Implementation of online problem based learning assisted by digital book with 3D animations to improve students' physics problem solving skills in magnetic field subject. *Journal of Technology and Science Education*, *12*(2), 379–396. https://doi.org/10.3926/jotse.1590

Putra, W. P., Sasmito, A., & Kass, R. D. D. (2021). The development of disaster mitigation curriculum to improve disaster response capabilities at the elementary school in Indonesia. *Eduvelop: Journal of English Education and Development*, *5*(1), 34–43. https://doi.org/10.31605/eduvelop.v5i1.920

Putri, F. I., Deswalman, D., & Tosin, A. I. (2024). Meta-analysis of PBL models on physics problem-solving abilities. *EduFisika*, *9*(1), 74–80. https://doi.org/10.59052/edufisika.v9i1.32648

Rofiah, N. H., Kawai, N., & Hayati, E. N. (2021). Key elements of disaster mitigation education in inclusive school setting in the Indonesian context. *Jàmbá: Journal of Disaster Risk Studies*, *13*(1), a1159. https://doi.org/10.4102/jamba.v13i1.1159

Rosana, D., Widodo, E., Setyaningsih, W., & Warno, D. S. (2020). Developing assessment instruments of PISA model to measure students' problem-solving skills and scientific literacy in junior high schools. *Jurnal Pendidikan Sains Indonesia*, *8*(2), 292–305. https://doi.org/10.24815/jpsi.v8i2.17468

Simamora, N. N., Istiyono, E., & Wilujeng, I. (2024). Development of an assessment instrument to measure high school students' collaboration skills in physics learning. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *10*(1), 39–46. https://doi.org/10.21009/1.10104

Sina, S. A., Uloli, R., & Abdjul, T. (2023). Website development as a physics learning media. *Jurnal Penelitian Pendidikan IPA*, *9*(8), 5874–5883. https://doi.org/10.29303/jppipa.v9i8.4189

Sugiarti, T., Kaniawati, I., & Aviyanti, L. (2017). Development of assessment instrument of critical thinking in physics at senior high school. *Journal of Physics: Conference Series*, *812*(1), 012018. https://doi.org/10.1088/1742-6596/812/1/012018

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sunarti, T., Satriawan, M., Prahani, B. K., & Rizki, I. A. (2024). Integrating digital books, 3D animations, and online problem-based learning to improve pre-service teachers' scientific literacy in general physics course. *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, *14*(2), 169–183. https://doi.org/10.26740/jpfa.v14n2.p169-183

Sung, Y. T., Chang, K. E., & Liu, T. C. (2022). The effects of integrating technology-supported learning on higher-order thinking. *Educational Technology & Society*, *25*(2), 1–14.

Suprapto, N., Rizki, I. A., & Cheng, T. (2024). Profile of students' physics critical thinking skills and prospect analysis of project-oriented problem-based learning model. *Journal of Educational and Social Research*, *14*(3), 134–149. https://doi.org/10.36941/jesr-2024-0062

Suryanti, N., & Nurhuda, N. (2021). The effect of problem-based learning with an analytical rubric on the development of students' critical thinking skills. *International Journal of Instruction*, *14*(2), 665–684. https://doi.org/10.29333/iji.2021.14237a

Suryawati, E., Osman, K., & Meerah, T. S. M. (2023). Applying PRISMA in educational research. *International Journal of Instruction*, *16*(2), 245–260. https://doi.org/10.29333/iji.2023.16214a

Ulu, Y., & Yerdelen-Damar, S. (2024). Metacognition and epistemic cognition in physics are related to physics identity through the mediation of physics self-efficacy. *Physical Review Physics Education Research*, *20*, 010130. https://doi.org/10.1103/PhysRevPhysEducRes.20.010130

Widana, I. W. (2017). Higher order thinking skills assessment (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *3*(1), 32–44. https://doi.org/10.21009/jisae.v3i1.4859

Wiyono, K., Sury, K., Hidayah, R. N., Nazhifah, N., Ismet, I., & Sudirman, S. (2022). STEM-based e-learning: Implementation and effect on communication and collaboration skills on wave topic. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *8*(2), 259–270. https://doi.org/10.21009/1.08208

Zakariya, Y. F. (2022). Cronbach's alpha in mathematics education research: Its appropriateness, overuse, and alternatives in estimating scale reliability. *Frontiers in Psychology*, *13*, 1074430. https://doi.org/10.3389/fpsyg.2022.1074430

***Sri Dewi (Corresponding Author)**
Physics Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang, Ketintang, Gayungan, Surabaya 60231, East Java, Indonesia.
Email: 24031635012@mhs.unesa.ac.id

Muhammad Satriawan
Physics Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang, Ketintang, Gayungan, Surabaya 60231, East Java, Indonesia.
Email: muhammadsatriawan@unesa.ac.id

Titin Sunarti
Physics Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang, Ketintang, Gayungan, Surabaya 60231, East Java, Indonesia.
Email: titinsunarti@unesa.ac.id

Iqbal Ainur Rizki
School of Education, Victoria University of Wellington
Level 8, Murphy Building, Kelburn Parade, Kelburn, Wellington 6012, New Zealand.
Email: iqbalainur19004@gmail.com