JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

Science Literacy Profile of Science Education Students in the Context of Applied Science: Analysis of Question Dimension Proportions and Score Achievements

Arvidhea Safira Gunawan[1]*, Firdha Yusmar[2], Siti Shofa Assyifa'ul Qulbi Barid[3], Rekfa Wika Amini[4], Fauziyatul Iffah[5], Sri Wahyuni[6]

[1,2,3,4,5,6] Department of Science Education, Faculty of Teaching and Education, University of Jember, Indonesia

*Email: arvidheasafira@unej.ac.id

ARTICLE INFO

Keywords:
Science literacy
Applied Science
Dimension
Score mapping

ABSTRACT

Purpose- This research was carried out to explain the scientific literacy profile of undergraduate science education students in the context of Applied Science, by examining (1) the proportion of science literacy question dimensions in multiple-choice test instruments, and (2) student achievement scores on each of the question dimensions of scientific literacy.

Methodology- This is a descriptive, quantitative study with a one-shot, cross-sectional measurement design. The study group included 115 pre-service science teachers from four parallel (A-D) classes, sampled using total sampling. The test contained 20 multiple-choice questions, which were pre-aligned to four dimensions of scientific literacy, including knowledge, attitudes, context, and competencies, and each of these dimensions was explicitly defined. Procedures included administering the test, scoring dichotomously (correct/incorrect), summarizing achievement by dimension using percent correct, and conducting Classical Test Theory item analysis by calculating percent correct for each item and classifying difficulty.

Finding- This study reveals that dimension achievement showed the highest performance in attitudes and moderate performance in knowledge, whereas performance in context and competencies was lower. Item difficulty analysis indicated that most items were easy ($p > 0.70$), a smaller portion were moderate (0.30–0.70), and none were difficult ($p < 0.30$), suggesting the instrument is effective for screening foundational mastery but may be less effective at differentiating mid-to-high performance levels.

Significance- This study provides an applied-science, explicit scientific literacy profile for science education students. It delivers practical diagnostic evidence to improve the MCQ instrument by balancing the blueprint, strengthening distractors, and increasing cognitive demand to capture contextual decision-making better.

Received 06 February 2026; Received in revised form 18 February 2026; Accepted 31 July 2026
Jurnal Eduscience (JES) Volume 13 No. 4 (2026)
Available online 30 August 2026

INTRODUCTION

The rapid growth of science and technology, along with the increasing flow of information in the digital world, makes scientific literacy ever more crucial for citizens who want to make evidence-based decisions about health, the environment, energy, and technology (Seema, 2024). The consensus report of the National Academies argues that the current information ecosystem facilitates the spread of scientific misinformation and suggests increasing the ability of the masses to evaluate scientific assertions, evidence, and sources (National Academies of Sciences Engineering & Medicine, 2025). However, scientific literacy has never had a single definitive definition. Abell (2013) identified two competing conceptions of scientific literacy: first, the idea that scientific literacy is knowledge about internal products and processes of science, facts, science procedures, and the logic of inquiry; and, the idea that scientific literacy is based on knowledge that can be applied in their personal, social, and civic lives in which citizens find themselves. Recent science education research builds on this second, more applied idea when it emphasizes that the ability to assess claims and distinguish credible scientific data has become a central element of science learning outcomes, particularly as students are exposed to digital material that blends science, opinion, and propaganda (Dabran-Zivan & Baram-Tsabari, 2025; Mellberg et al., 2025; Wahyuni et al., 2026). This distinction matters for the present study because it signals that scientific literacy cannot be reduced to the amount of content knowledge a person holds; it also concerns whether that knowledge can be mobilized to reason and act.

The improvement of scientific literacy at the global level is also considered an important indicator in international measures. The Organization for Economic Co-operation and Development (OECD) (OECD, 2023b) defines scientific literacy as the ability to engage with science-related problems, utilize scientific knowledge and procedures to develop informed solutions, and analyze evidence across various life scenarios. The science literacy framework presented in PISA 2025 outlines the scientific literacy in three aspects: (a) scientific competence, (b) scientific knowledge, including content, procedural, and epistemic aspects, and (c) context, on a personal, local/national, to global range (OECD, 2023c). Read against Abell (2013)The PISA framework can be viewed as a way to operationalize the second idea of science literacy because it explicitly calls for content knowledge to be accompanied by the ability to apply it in a context, rather than being measured in isolation. As a result, scientific literacy goes beyond the ability to master concepts and encompasses data management, investigation design, and evaluation, and the critical analysis of the quality of evidence.

In Indonesia, the need to strengthen scientific literacy is becoming increasingly apparent, since the PISA score has consistently been below the OECD average. According to data from PISA 2022, Indonesians' average science score was 383 points, which is significantly lower than the OECD average (485 points) (OECD, 2023a). This is not a one-time shortfall: OECD's own trend analysis identifies the drop in Indonesia's science performance between the 2015 and 2018 assessment cycles as one of the steepest declines recorded among all participating countries and economies, and the 2022 results remained essentially flat relative to 2018, rather than showing any sign of recovery (OECD, 2023a, 2023b). In addition, the Indonesian PISA report notes that the proportion of students who achieve at least Level 2 in science is significantly lower than the OECD average (Ayu et al., 2025). These results suggest that to enhance scientific literacy, it is necessary to systematically use pedagogies and assessments grounded in the dimensions of scientific literacy.

This urgency carries particular weight in higher education, particularly in science education, where the scientific literacy of prospective teachers plays a strategic role (Sarini et al., 2024). Science education students are training to become the very people responsible for cultivating scientific literacy in the next generation;0 if their own literacy remains narrowly conceptual, that limitation is likely to be reproduced in their future classrooms. As prospective teachers, they are expected not merely to understand concepts but also to model scientific thinking, assess evidence, and relate science to decision-making in personal, social, and global contexts (Ceyhan et al., 2021). The science education research framework positions scientific literacy as the primary goal of modern science education because it is directly linked to inquiry-based teaching, scientific argumentation, and information evaluation (Tirumalamba et al., 2024). Thus, the scientific literacy of science education students in the context of applied science is key to developing the professional competencies of prospective teachers who can connect science to real-world problems.

However, studies of scientific literacy in Indonesia remain quite diverse and primarily focus on aggregate scores or broad categories, often using questionnaires or composite indicators. (Yanto et al., 2025). As a result, these investigations are not granular enough to determine which specific dimensions are evaluated and which still cause performance issues for the participant. Such deficiencies include the interaction between competence, knowledge, and attitude (Yudhistian & Wulandari, 2025). Nor is this purely a conceptual gap: comparative analyses of assessment construction itself, such as how many items in an instrument are actually devoted to each dimension of scientific literacy, are markedly sparser than research on learning interventions, even though PISA stresses that literacy performance is always elicited contextually by the specific demands of the items used to measure it, so that the quality and composition of those items fundamentally shapes what is being measured in the first place (OECD, 2023c). Thus far, few studies appear to have combined an item-level mapping of dimension proportions with a corresponding dimension-level analysis of student achievement in a single applied science course for pre-service science teachers, which is precisely the gap the present study is designed to close.

This gap is consequential because assessment plays a double role: it is simultaneously an instrument for quantifying what students have achieved and a catalyst that shapes what and how they subsequently learn (Schüttpelz-Brauns, 2022). The present study addresses this gap within the Fundamentals of Science course, a foundational applied science course in the Science Education curriculum, whose formally stated learning outcome is that students can implement the scientific method as a structured, stepwise process for scientific inquiry to solve problems and generate knowledge. Because this learning outcome is defined in explicitly procedural and applied terms rather than in terms of content recall alone, the course provides a naturally appropriate site for examining scientific literacy as a multidimensional construct: an evaluation instrument dominated by content-knowledge items would, by construction, fail to capture whether students have actually achieved the course's own stated outcome. When items of content knowledge dominate an evaluation instrument in such a course, the resulting outcomes tend to over-weight conceptual mastery at the expense of higher-order cognitive processes, data reasoning, evidentiary evaluation, or the design of inquiry, which are the very substance of scientific literacy and of this course's learning outcome. Conversely, a test that maintains a deliberate balance of items across the knowledge, attitude, context, and competency dimensions can yield results that provide a considerably more precise diagnostic map for improving both the instrument and the course itself (Dauer et al., 2025). International research on instrument development also highlights the significance of competence, often considered the cornerstone of scientific literacy, and of scientific knowledge and attitudes/identity, in the ability to demonstrate competence in authentic contexts. (Zhang et al., 2023).

Based on these conditions, the current study aims to characterize the scientific literacy profile of Science Education students in the context of Applied Science by two main products: (1) science literacy question dimensions proportion in multiple-choice test instruments, and (2) student achievement scores on each of the question dimensions of scientific literacy. The novelty of this study is in the deliberate linking of the analysis of assessment content (how scientific literacy dimensions were represented in the test items) to the analysis of student performance (how students scored on each of the scientific literacy dimensions) in the same, but real, assessment, in such a way that the analysis of assessment content becomes a basis for the analysis of student performance and vice versa, thus generating an actionable diagnostic for two audiences at once: those who design the item bank and those who design the learning experiences of the Fundamental of Science course. This methodology ensures that the scientific literacy profile goes beyond aggregate scoring to provide a dimensional mapping of related achievements, which can be used to make targeted recommendations to improve underdeveloped aspects of scientific literacy.

The research questions for this study are: (1) What percentage of the scientific literacy dimension (knowledge, attitude, context, and competency) covered in the scientific literacy test that was used are included in the multiple-choice test of the Fundamentals of Science course? (2) What is the achievement level of Science Education students on each of these scientific literacy dimensions, and what is the pattern of relative strengths and weaknesses in the students' achievement profile?

METHODOLOGY

Research Design

This study adopted a descriptive, quantitative research design with a one-shot, cross-sectional measurement methodology. The assessments were conducted at a single point using a multiple-choice quiz that mapped participants' science literacy profiles in the Applied Science context. (Zuleika & Legiran, 2022). The selected methodology was intended to (1) define the composition of science-literacy dimensions captured in the test items and (2) report the scores of the students on science literacy dimensions.

The research was conducted in the Fundamentals of Science course, within the context of applied science, the foundational curricular component of the Science Education Study Program. The procedures and measurement instruments were standardized across all participants (same items, time, completion requirements, and scoring schema) to ensure that the outcomes reflected equal achievement.

Participant

The sampling method used was total sampling, meaning that all students enrolled in the four parallel classes of the Fundamentals of Science Course (Classes A to D) during the semester in which data collection took place were invited, rather than being narrowed through random or purposive selection. This approach was deliberately chosen to minimize partial selection bias and to preserve the natural composition of the course as it is actually taught, ensuring the resulting literacy profile reflects the full student population in that offering rather than an artificially restricted sample. Of the 120 students initially enrolled across the four classes, 115 students met all inclusion criteria. They provided complete, usable responses after the data-cleaning process described below, yielding final class sizes of 29 students in Class A, 28 in Class B, 29 in Class C, and 29 in Class D. Exclusion criteria were: (a) response sheets in which the student did not complete twenty items, since partial completion would distort the calculation of dimension-specific scores and item-difficulty indices; and (b) response patterns indicating non-serious engagement, such as identical answers selected across all items regardless of content, identified through initial screening of the raw data matrix.

Data Collection

The data collection process commenced by scheduling and standardizing the administration of the tests to all the participants under identical conditions. Time limit, spelled-out guidelines, and uniform completion and scoring guidelines were put in place. The multiple-choice question (MCQ) test was given in a single sitting after the lecture via Wayground to maintain the integrity of item content, instructions, and time constraints.

Based on the post-assessment, the responses were shipped to a matrix database (rows are students; columns are items). Data completeness was ensured, and cleaning was performed using pre-established exclusion criteria, yielding a valid dataset as the final sample. At the same time, items were mapped to dimensions at the item level to clearly record the proportion of each dimension and the related achievement. Coding guidelines that provide operational definitions and indicators for each dimension were developed. The content validity was assessed using the best science learning assessment scholars, who found that the instrument is acceptable and valid in its item-dimension classification

Instrument

The instrument used was a 20-item multiple-choice (MCQ) test with a single correct answer, designed to reflect four dimensions of scientific literacy. Table 1 summarizes the sub-aspects covered under each dimension, which served as the basis for item writing. These dimensions are consistent with the modern understanding of science literacy, which emphasizes knowledge integration, competence in scientific processes, and their applicability to real-world situations (OECD, 2023c). Assigning a rating of one point to a correct answer and zero to an incorrect answer, the total score range was 0-20. The derivation of dimension-specific scores involved item grouping.

Table 1. Scientific Literacy Dimensions and Their Sub-Aspects covered by the Instrument

Dimension	Sub-aspects/indicators covered
Knowledge	Content knowledge, procedural knowledge, epistemic knowledge
Attitude	Curiosity, openness to evidence, precision/carefulness, critical stance toward information sources, and environmental consciousness
Context	Personal, social, global
Competency	Explanation of phenomena, assessment/design of investigations, interpretation of data

Data Analysis

Data analysis was done at various levels. At the item level, the frequency (n) and proportion (percentage) of items in each dimension were computed and presented in a distribution table and as bar charts. Mean dimension-specific and standard deviation, as well as percentage correct, had been calculated at the student level (correct responses/responses per dimension). The item difficulty statistics (p-values) were generated to identify the hardest and easiest items in each dimension. The categories for item difficulty (p-value) levels based on the guidelines Rashwan et al. (2024) are shown in Table 2. All the processing was done in SPSS 26 and Microsoft Excel.

Table 2. Item Difficulty Categories Based on p-value

Category	p-value range
Difficult	$p < 0.30$
Moderate	$0.30 \leq p \leq 0.70$
Easy	$p > 0.70$

FINDINGS

Data Screening of Participants

According to science literacy test results from four classes in which the test was administered, 115 of 120 students passed and met the inclusion criteria. Incomplete data were excluded, and the resulting sample sizes after data cleansing are given in Table 3.

Table 3. Overall Test Participant in All Classes

Class	N (initial)	N (final)
A	30	29
B	30	28
C	30	29
D	30	29

Mapping of Scientific Literacy Dimensions

The analysis of score achievement precedes the classification of each test item into distinct scientific literacy dimensions. Figure 1 shows the percentage of items to be given to each dimension in the multiple-choice questionnaire. Items 1-5 were coded as the knowledge dimension (content, procedural, and epistemic); items 6-10 as the scientific attitudes dimension; items 11-15 as the context dimension (personal, social, and global); and items 16-20 as the competency dimension (explaining scientific phenomena, designing/evaluating investigations, and interpreting scientific data and evidence). Table 3 lists the dimensions of scientific literacy mapped to each test item and achievement indicator.

As shown in Figure 1, the 20 items were distributed evenly across the four dimensions, with 5 items (25%) assigned to each dimension. This balanced proportion was an intentional feature of the test blueprint, designed to ensure that no single dimension would be over-represented at the expense of the others, so that the achievement profile discussed in the following section can be meaningfully compared across dimensions without the confound of unequal item weighting.

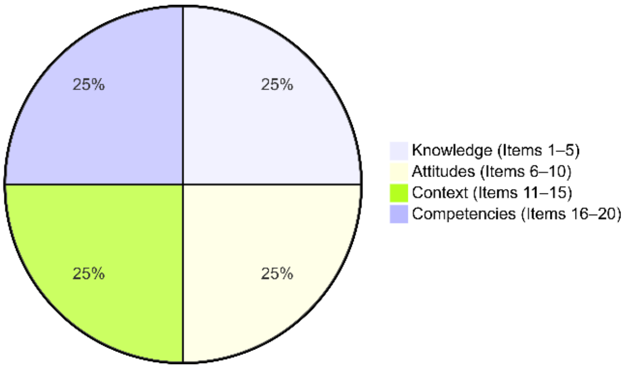

Figure 1. The Proportions of Scientific Literacy Dimensions in the MCQ Test

To make this mapping transparent and traceable to the underlying construct, Table 4 presents the complete correspondence between each item number, its assigned scientific-literacy dimension, and the specific indicator that the item was designed to measure, allowing readers to verify the alignment between item content and the intended dimension before proceeding to the achievement results reported in the following sections.

Table 4. Mapping Science Literacy Dimensions in Items

Item	Dimenssion	Indicator
1–2, 5	Knowledge (Content)	Identifying and applying core scientific concepts in everyday contexts
3	Knowledge (Procedural)	Understanding experimental procedures to test hypotheses
4	Knowledge (Epistemic)	Understanding how scientific knowledge changes with new evidence
6–10	Attitudes	Curiosity, openness to evidence, carefulness, critical source checking, and environmental concern
11	Context (General)	Recognizing science-in-context as real-life situations where science is used
12–13	Context (Personal)	Using science-related information for personal health/choices
14	Context (Social)	Connecting science to societal issues and public decision-making
15	Context (Global)	Connecting science to global-scale issues (e.g., energy, environment)
16–20	Competencies	Explaining phenomena, designing/evaluating investigations, interpreting data and evidence

Total Score Achievement

In general, students' total scores on the science literacy test provide an initial overview of their overall mastery level before further analysis of each dimension. Of the 20 MCQ items, students' scores across 4 classes (M = 16.68, SD = 2.50). which is then converted to a 0–100 point scale to facilitate interpretation (score 0–20 multiplied by 5); thus, the score is equivalent (M = 83.40, SD = 12.54) points on a 0–100 scale, with a score range of 82–86. A descriptive statistics summary of the total score across 4 classes, including mean and standard deviation values, is presented in Table 5.

Table 5. Descriptive Statistics of Total Scores by Class

Class	Mean ± SD	Min–Max
A	81.55 ± 11.89	55–100
B	81.95 ± 14.74	50–100
C	84.50 ± 11.37	60–100
D	85.70 ± 12.16	60–100
Overall	83.40 ± 12.54	50–100

While Table 5 summarizes the central tendency and spread of scores numerically, it does not on its own reveal potential asymmetries, outliers, or overlapping performance ranges between classes. Figure 2, therefore, complements this table by visualizing the same data as a boxplot, allowing the median position, interquartile range, and any extreme values within each class to be inspected directly rather than inferred from a mean and standard deviation alone.

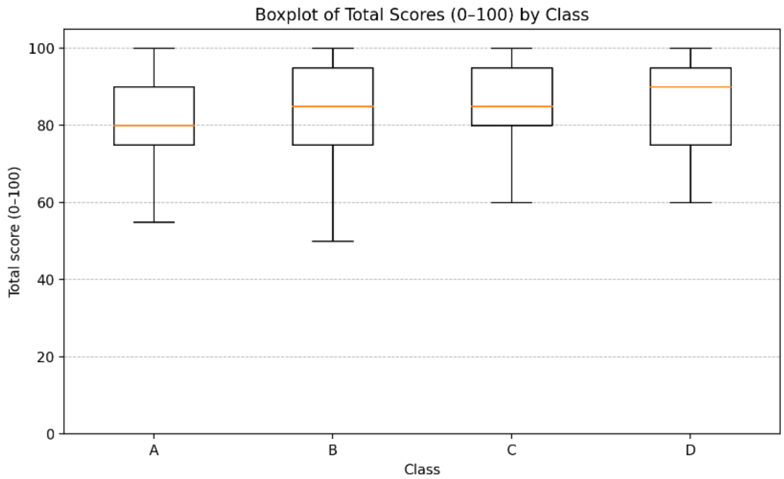

Figure 2. Boxplot Diagram of the Distribution of Scores by Class

Achievement Scores on The Scientific Literacy Dimension

The achievement scores for each dimension show contrasting patterns across the aspects of scientific literacy. The outcome shows that students scored highest on the attitude dimension, indicating positive traits such as curiosity, openness to evidence, thoroughness, and caution when using information sources. Achievement in the knowledge dimension ranked medium-high, indicating that most students had a fairly good command of conceptual basics and reaction/epistemic features. In its turn, success in the context and competency dimensions was the last, which presupposes that the main difficulty of students is the use of science in their personal, social, and global context and in addressing the needs of the scientific process, especially the interpretation of data/evidence and the design/evaluation of investigations. The mean, standard deviation, and percentage of correct responses per dimension are presented as summary statistics in Table 6.

Table 6. Descriptive Statistics of Students' Scientific Literacy Achievement by Dimension

Dimension	Items	Mean ± SD	Percent correct (%)
Knowledge	1–5	3.85 ± 1.19	77.0
Attitudes	6–10	4.51 ± 0.69	90.3
Context	11–15	1.50 ± 0.58	30.1
Competencies	16–20	1.83 ± 0.40	36.5

Beyond the summary values reported in Table 6, the relative size of the achievement gap between dimensions is easier to grasp visually than numerically. Figure 3, therefore, presents the same percentage-correct values in a bar-chart format, contrasting the comparatively strong performance in the attitude dimension and the comparatively weak performance in the context and competency dimensions immediately apparent.

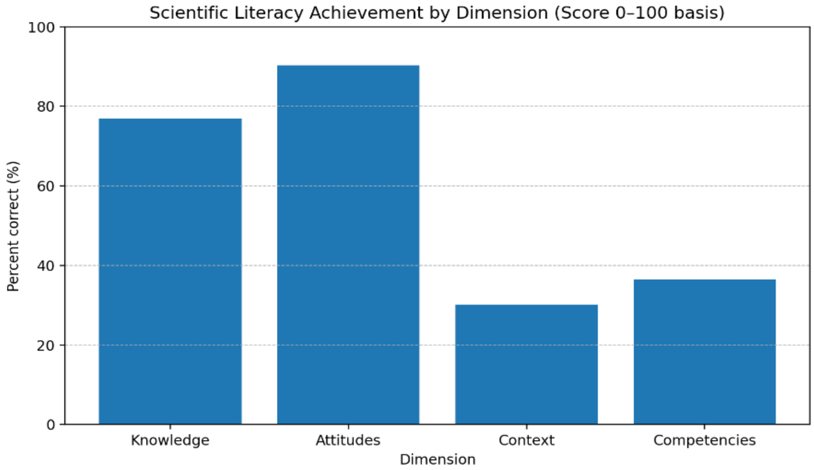

Figure 3. Scientific Literacy Achievement by Dimension

Item Difficulty Analysis and Instrument Diagnosis

The results of the item difficulty analysis indicate that the instrument's difficulty level is concentrated in the easy category, with only a small portion in the moderate category and none in the very difficult category. When viewed by dimension, the items in the attitude dimension are generally in the easy category. In contrast, the most challenging items tend to appear in dimensions that require application and reasoning (e.g., context and competence), and some in the knowledge dimension, which requires a more precise understanding of concepts. A summary of the p-proportion correct (p-value) per item, along with the dimensions, difficulty categories, and summary indicators, is presented in Table 7 as a basis for diagnosis.

Table 7. Item Difficulty (p-value) and Instrument Diagnosis

Item	Dimensions	p-value	Percent correct (%)	Category
1	Knowledge	0.896	89.6	Easy
2	Knowledge	0.765	76.5	Easy
3	Knowledge	0.887	88.7	Easy
4	Knowledge	0.713	71.3	Easy
5	Knowledge	0.591	59.1	Moderate
6	Attitudes	0.757	75.7	Easy
7	Attitudes	0.870	87.0	Easy
8	Attitudes	0.878	87.8	Easy
9	Attitudes	0.904	90.4	Easy
10	Attitudes	0.939	93.9	Easy
11	Context	0.809	80.9	Easy
12	Context	0.765	76.5	Easy
13	Context	0.609	60.9	Moderate
14	Context	0.783	78.3	Easy
15	Context	0.791	79.1	Easy
16	Competencies	0.826	82.6	Easy
17	Competencies	0.783	78.3	Easy
18	Competencies	0.870	87.0	Easy
19	Competencies	0.696	69.6	Moderate
20	Competencies	0.826	82.6	Easy

DISCUSSION

Based on the research findings presented, MCQs are well-suited to efficiently mapping cognitive achievement in large classes, especially when the items are structured around a clear blueprint and defined indicators. (Abdellatif, 2023). The dataset consisted of 115 respondents from four classes (A–D), each completing the same 20 items, which is an adequate sample size for descriptive analysis and for an initial diagnosis of instrument quality using classical test theory (CTT) (Fauzie et al., 2021; Gebremichael et al., 2025). Two features make the dataset well-suited for this diagnosis. First, because every respondent answered the same 20 items, the risk of missing data at the respondent level is low, allowing dimension scores to be interpreted directly without estimating missing values. Second, the data show, for each item, how many students answered correctly and how many did not; this is the basic information needed to calculate each item's difficulty index (p-value) under CTT and to check whether item difficulty is appropriately distributed for diagnostic purposes (Ohiri & Okoye, 2023; Rashwan et al., 2024).

This study thus directly addresses the gap mentioned in the introduction. Most previous studies of scientific literacy in Indonesia report only a single overall score, which does not show how performance differs across the different aspects of scientific literacy. This study looked at both sides of the picture. First, mapping the 20 MCQ items showed they were evenly distributed across the four dimensions, with 5 items per dimension. Second, once achievement was broken down by dimension, a very different picture appeared: performance was considerably higher in two of the four dimensions than in the other two. This is exactly why combining content analysis (how the instrument is built) with performance analysis (how students actually performed) matters. If this study had reported only a single total score, as earlier Indonesian studies on scientific literacy typically do, the headline figure for these 115 students would have been a moderate-to-high average ($M = 83.40$ on a 0–100 scale; see Table 5). That single number would have hidden the unevenness across dimensions rather than revealing it.

After disaggregating achievement by dimension, a clear and consistent pattern emerges. The students' results were best in the attitude dimension (90.3% correct), followed by knowledge (77.0%), competency (36.5%), and context (30.1%). This pattern is not entirely consistent with purely content-focused models of scientific literacy, which predict that once students master core knowledge, reflected in their relatively high knowledge scores, that knowledge is integrated into personal, social, and global contexts and mobilized into skills such as interpreting data or designing investigations. Instead, this transfer clearly does not occur at the same rate. This is in line with the science literacy framework for PISA 2025, which sets forth content, procedural and epistemic knowledge as essential but not sufficient elements in the scientific competence, with a particular emphasis on the epistemic and procedural knowledge, specifically knowledge about how science works (investigative practice, the use of evidence, the justification of claims), which are consistently more difficult to demonstrate than factual content knowledge (Lin & Puntambekar, 2025; Zetterqvist & Bach, 2023). Similar findings have also been reported in studies of prospective teachers, where understanding of science literacy and dimensions of knowledge related to the nature of science (NOS) and the use of evidence often require systematic reinforcement in lectures (Al Sultan et al., 2021; Nieminen & Ketonen, 2024). In studies of science literacy among prospective teachers in Indonesia, varying achievement across aspects is also commonly observed and is often influenced by instrument characteristics, particularly the context and the demands of data-based reasoning (Himmah et al., 2024; Yudhistian & Wulandari, 2025). The relative underachievement in the context dimension is also found in the science literature, which suggests that the ability to recognize a concept in a text is not equivalent to the ability to recognize how that concept applies in an unfamiliar, real-world context (Cian, 2020). The even lower achievement on competency items supports this, as data interpretation, evidence evaluation, and investigation design are advanced skills that require integrating knowledge with reasoning, not just recognition (Janoušková et al., 2023).

Item difficulties provide information about the causes of the differential achievement pattern. 19 of the 20 items were in the easy range ($p > 0.70$), 3 were in the moderate range, and none were in the difficult range. In fact, among the items that scored in the moderate range, one was from each of the dimensions of knowledge, context, and competency, and none of the five attitude items fell in the moderate range; all five were easy. This distribution indicates that the overrepresentation of easy items is not evenly distributed across dimensions, but is particularly high in the attitude dimension, in which multiple-choice items might be more vulnerable to social-desirability cueing, as students who do not have a well-developed disposition towards a given construct might be able to select the correct answer by recognizing the "correct-sounding" choice rather than by internalizing the item's attitude. (García-Carmona, 2025; Hu & Bi, 2025). However, the three dimensions of knowledge, context, and competency each had at least one moderately difficult item; it is specifically two of these three dimensions (context and competency) that achievement was also relatively low, likely one reason why they were also more likely to be classified as difficult, as items requiring procedural or epistemic reasoning, contextual transfer, or data-based competency are inherently harder to answer through recognition alone than items that simply require recalling a fact, and this same underlying demand is what makes such items both harder to classify as easy and harder to answer correctly (Lee et al., 2024; Rame & Hoinbala, 2024; Srisomsak et al., 2025). In addition to the dimension-specific explanation, the overall pattern of easier items on the instrument is also consistent with a common pattern of easy item types overall in the classroom level of assessment, where poorly functioning distractors, or items pitched mainly at the recognition level, tend to inflate the proportion of easy items on an instrument (Chauhan et al., 2023; Hartono et al., 2023).

The distribution of the p-values suggests that the instrument has a good ability to function as a screening tool for foundational mastery, as nearly all items are sufficiently easy for most students to achieve, but the instrument is not yet well-suited to distinguish between students with medium-to-high performance levels, where the real differences in scientific literacy are likely to occur (Nur et al., 2025; Tani & Nchia, 2025). The lack of difficult items is itself informative: it indicates that even the context and competency dimensions, which are conceptually the more demanding dimensions of scientific literacy, were not difficult enough to create a distinction between high and low performance levels, a finding that is associated with the need for rebalancing the blueprint to include more items that demand higher reasoning rather than greater topical difficulty (Abdellatif, 2023; Pawade et al., 2022).

These results should be interpreted in light of the intrinsic limitations of the multiple-choice test format used to measure scientific literacy. At best, an MCQ will be able to assess whether a student can choose a scientifically appropriate answer from a closed set of answers; it will not assess whether the student can form an evidence-based argument without prompting; it will not assess whether the student can design an original investigation from scratch; or it will not assess whether the student can justify a claim in their own words, which are all part of the competency dimension as defined in this study and in the PISA framework more generally (OECD, 2023c). This may help explain why competency achievement remained comparatively low even though every competency item was classified as easy or moderate in difficulty: a multiple-choice format can allow students to select a correct answer through partial cues, elimination techniques, or fortunate guessing without genuinely possessing the underlying competency, meaning the true competency gap may in fact be larger, not smaller, than the reported 36.5% suggests. The same caveat applies to the attitude dimension, and even more so, because the items on the attitude scale in an MCQ format require students to recognize or endorse a scientifically appropriate attitude, but not necessarily to demonstrate an attitude in an authentic decision, the 90.3% achievement percentage is likely an overestimate of the extent to which such attitudes are internalized in practice. Overall, these considerations suggest that the diagnostic value of this instrument lies less in its absolute achievement percentages and more in the relative patterns it reveals across dimensions, particularly for the context and competency dimensions, where the multiple-choice format has the greatest impact on what is actually measured; and that future versions of the instrument (and associated course it serves) would benefit from having constructed response or performance based tasks specifically for the context and competency dimensions (Arfiani et al., 2025; Irawan et al., 2025; Steinbach et al., 2025).

The results have clear instructional implications for the Fundamentals of Science course in particular. The relatively low levels of achievement in these two dimensions (context and competency) suggest that students' ability to use the scientific method as a structured approach to problem-solving and knowledge creation is not fully realized, contrary to the course's learning outcomes. Students are seen as more competent in recognizing scientifically appropriate knowledge and attitudes than they are in acting upon that knowledge in the structured and evidence-based thinking that the scientific method requires (Çalik & Wiyarsi, 2025; Högström et al., 2025; Hüfner et al., 2025). Two clear ways forward for improving the course are outlined here. In the first place, learning activities could be restructured to allow students to have more intentional practice in applying their knowledge to new personal, social, and global contexts, such as through tasks related to socio-scientific issues, which the literature shows can be used to strengthen context-based reasoning (Rathnayake & Senevirathne, 2024; Zhang et al., 2023). Second, because competency items proved the most challenging of all, structured opportunities for students to practice data interpretation and to design or critique investigations (preferably hands-on or project-based, rather than only covered in lecture format), would target the dimension most closely related to the procedural learning outcome (Khery et al., 2022). The overall message in both directions is the same: improving scores on this instrument is not an end in itself; it is a diagnostic tool for redesigning how the scientific method is taught and assessed in the Fundamentals of Science course.

Taken together, these findings point to this study's dual contribution: item-level mapping of how scientific literacy dimensions are distributed in real course assessments, combined with dimension-level analysis of student achievement on the same assessments, thus helping to close some of the gaps identified at the beginning of this paper. The study concretely demonstrates why aggregate literacy scores are incomplete diagnostics: achievement is strong in the attitudes and knowledge dimensions but much weaker in context and competence, a pattern that item-level analysis traces partly to the near-total dominance of easy items, particularly in the attitudes dimension, and partly to the significantly higher reasoning demands of context- and competence-based items. The instrument itself is adequate for screening for basic mastery but is not yet sensitive enough to distinguish between stronger students or to capture higher-order competencies; if anything, its multiple-choice format likely underestimates the true competency gap while overestimating how internalized students' scientific attitudes actually are. These conclusions chart two complementary directions for improvement: rebalancing the instrument design toward items with higher genuine reasoning demands, supplemented with performance-based tasks for both context and competency dimensions, and redesigning the Science Foundations course itself so that improving test scores is not the primary goal, but rather a means toward more deliberate practice in context-based and evidence-based scientific reasoning (Dewi et al., 2023).

CONCLUSION

Based on the discussion, it is known that this study has mapped the science literacy profile of science education students in the context of applied science through (1) analysis of the composition of science literacy dimensions in a 20-item MCQ instrument and (2) analysis of student performance achievements per dimension, while also conducting an initial diagnosis of the quality of CTT-based instruments. The dataset, comprising 115 students with complete responses to all items, was deemed suitable for descriptive analysis and item diagnosis. It showed a pattern of achievement that was relatively stronger in the attitude dimension than in other dimensions. In contrast, the knowledge dimension and indicators that required transfer to authentic contexts and scientific processes (context-competence) still needed strengthening. In terms of instrument quality, the distribution of the difficulty index (p-value) shows a dominance of easy items, with a small proportion of moderate items and none in the difficult range, making the instrument effective for screening basic mastery but not yet optimal for distinguishing abilities in the medium-high range. Recommendations aim to balance the blueprint, increase demand for evidence-based reasoning, and improve the quality of distractors in overly easy items, thereby making the instrument more diagnostic and better aligned with the science literacy framework.

ACKNOWLEDGEMENT

The author would like to thank the Institute for Research and Community Service (LP2M) at the University of Jember for its financial support through the Early-Career Lecturer Research Grant Program, which was funded from the University of Jember's internal budget for Fiscal Year 2026, in accordance with Agreement No. 3944/DST/UN25.C1/LT/2026.

REFERENCES

Abdellatif, H. (2023). Test results with and without blueprinting: Psychometric analysis using the Rasch model. *Educación Médica*, *24*(3), 100802. https://doi.org/10.1016/j.edumed.2023.100802

Abell, S. K. (2013). *Handbook of Research on Science Education*. Routledge. https://doi.org/10.4324/9780203824696

Al Sultan, A., Henson, H., & Lickteig, D. (2021). Assessing preservice elementary teachers' conceptual understanding of scientific literacy. *Teaching and Teacher Education*, *102*, 103327. https://doi.org/10.1016/j.tate.2021.103327

Arfiani, Y., Yulianti, D., Masturi, M., & Nugroho, S. E. (2025). Feasibility analysis of instruments for scientific reasoning skills for prospective science teachers using the Rasch model. *Unnes Science Education Journal*, *14*(2), 405–416. https://doi.org/10.15294/usej.v14i2.25493

Ayu, G. N., Putri, C. A., Riyanto, A. R., & Koto, I. (2025). The Scientific Literacy Competence of Students in Indonesia and Mexico Based on PISA 2022: An International Comparative Study. *TOFEDU: The Future of Education Journal*, *4*(5), 1033–1038. https://doi.org/10.61445/tofedu.v4i5.525

Çalik, M., & Wiyarsi, A. (2025). The effect of socio-scientific issues-based intervention studies on scientific literacy: a meta-analysis study. *International Journal of Science Education*, *47*(3), 399–421. https://doi.org/10.1080/09500693.2024.2325382

Ceyhan, G. D., Lombardi, D., & Saribas, D. (2021). Probing into Pre-Service Science Teachers' Practices of Scientific Evaluation and Decision-Making on Socio-scientific Issues. *Journal of Science Teacher Education*, *32*(8), 865–889. https://doi.org/10.1080/1046560X.2021.1894762

Chauhan, G. R., Chauhan, B. R., Vaza, J. V, & Chauhan, P. R. (2023). Relationships of the number of functioning distractors with the item difficulty index and the item discrimination power in the multiple-choice questions. *Cureus*. https://doi.org/10.7759/cureus.42492

Cian, H. (2020). The influence of context: comparing high school students' socioscientific reasoning by socioscientific topic. *International Journal of Science Education*, *42*(9), 1503–1521. https://doi.org/10.1080/09500693.2020.1767316

Dabran-Zivan, S., & Baram-Tsabari, A. (2025). The Importance of Science Education, Scientific Knowledge, and Evaluation Strategies for the Successful Detection of COVID-19 Misinformation. *Science Education*, *109*(6), 1733–1759. https://doi.org/https://doi.org/10.1002/sce.70000

Dauer, J., Gaona, V. B., & Meulenbroeks, R. (2025). Vision II Scientific Literacy Assessment Instruments for Secondary Education: A Scoping Review. *International Journal of Education in Mathematics, Science and Technology*.

Dewi, H. H., Damio, S. M., & Sukarno, S. (2023). Item analysis of reading comprehension questions for the English proficiency test using the Rasch model. *REID (Research and Evaluation in Education)*, *9*(1), 24–36. https://doi.org/10.21831/reid.v9i1.53514

Fauzie, M., Pada, A. U. T., & Supriatno, S. (2021). Analysis of the difficulty index of the item bank according to cognitive aspects during the Covid-19 pandemic. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *25*(2). https://doi.org/10.21831/pep.v25i2.42603

García-Carmona, A. (2025). Scientific Thinking and Critical Thinking in Science Education. *Science & Education*, *34*(1), 227–245. https://doi.org/10.1007/s11191-023-00460-5

Gebremichael, M. W., Baraki, B., Mehari, M.-A., & Assalfew, B. (2025). Item analysis of multiple choice questions from the assessment of health sciences students, Tigray, Ethiopia. *BMC Medical Education*, *25*(1), 441. https://doi.org/10.1186/s12909-025-06904-6

Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2023). Exploration of diagnostic testing instruments: Validity, reliability, and item characteristics. *Pegem Journal of Education and Instruction*, *13*(3). https://doi.org/10.47750/pegegog.13.03.39

Himmah, R., Jatmiko, B., Prahani, B. K., & Rizki, I. A. (2024). Science Literacy Competency Profile of Science

Education Students in Understanding the Concept of Thermodynamics. *IJORER : International Journal of Recent Educational Research*, *5*(6), 1600–1615. https://doi.org/10.46245/ijorer.v5i6.724

Högström, P., Gericke, N., Wallin, J., & Bergman, E. (2025). Teaching Socioscientific Issues: A Systematic Review. *Science & Education*, *34*(5), 3079–3122. https://doi.org/10.1007/s11191-024-00542-y

Hu, X., & Bi, H. (2025). Exploring and validating the componential model of students' scientific critical thinking in science education. *Thinking Skills and Creativity*, *55*, 101695. https://doi.org/10.1016/j.tsc.2024.101695

Hüfner, S., Weirauch, K., List, F., Menthe, J., & Abels, S. (2025). Context-based science education to promote diversity, equity, and inclusion – a systematic literature review on the understanding of context in science education. *Studies in Science Education*, 1–41. https://doi.org/10.1080/03057267.2025.2563946

Irawan, I. D. A., Indraloka, R. M., Basri, N. A., Salmah, U., & Parno, P. (2025). Analysis of concept-understanding test items on static fluid material using the Rasch model. *Jurnal Pendidikan Fisika*, *13*(1), 1–13. https://doi.org/10.26618/jpf.v13i1.15687

Janoušková, S., Pyskatá Rathouská, L., Žák, V., & Urválková, E. S. (2023). The scientific thinking and reasoning framework and its applicability to manufacturing and services firms in the natural sciences. *Research in Science & Technological Education*, *41*(2), 653–674. https://doi.org/10.1080/02635143.2021.1928048

Khery, Y., Sarjan, M., Ahzan, S., & Efendi, I. (2022). Konseptualisasi literasi sains mengacu pada kerangka Sains Pisa sejak tahun 2000. *Educatoria : Jurnal Ilmiah Ilmu Pendidikan*, *2*(4), 200–231. https://doi.org/10.36312/ejiip.v2i4.117

Lee, H. Y., Yune, S. J., Lee, S. Y., Im, S., & Kam, B. S. (2024). The impact of repeated item development training on the prediction of medical faculty members' item difficulty index. *BMC Medical Education*, *24*(1), 599. https://doi.org/10.1186/s12909-024-05577-x

Lin, F., & Puntambekar, S. (2025). Epistemic scaffolding: understanding and designing the support for epistemic growth in science. *International Journal of Science Education*, *47*(6), 769–793. https://doi.org/10.1080/09500693.2024.2349973

Mellberg, S., Danielsson Friberg, A., & Nygren, T. (2025). Science education against misinformation: an educational intervention in upper-secondary schools. *International Journal of Science Education*, 1–33. https://doi.org/10.1080/09500693.2025.2571131

National Academies of Sciences, Engineering, and Medicine. (2025). *Understanding and Addressing Misinformation About Science* (K. Viswanath, T. E. Taylor, & H. G. Rhodes (eds.)). The National Academies Press. https://doi.org/10.17226/27894

Nieminen, J. H., & Ketonen, L. (2024). Epistemic agency: a link between assessment, knowledge, and society. *Higher Education*, *88*(2), 777–794. https://doi.org/10.1007/s10734-023-01142-5

Nur, F., Makmur, F., & Fajriah, N. A. (2025). Test instrument development with local-wisdom context, assisted by the iSpring application, in the mathematics course. *Matematika Dan Pembelajaran*, *13*(1), 43–63. https://doi.org/10.33477/mp.v13i1.9225

OECD. (2023a). *Education GPS: Indonesia – Student performance (PISA 2022)*.

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

OECD. (2023c). *PISA 2025 science framework*.

Ohiri, S. C., & Okoye, R. O. (2023). Application of Classical Test Theory as Linear Modeling to Test Item Development and Analysis. *International Research Journal of Modernization in Engineering Technology and Science*, *5*(10), 2152–2159. https://doi.org/10.56726/IRJMETS45379

Pawade, Y. R., Chalak, A. S., & Pawade, D. Y. (2022). Blueprints in biochemistry: Systematic assessment in undergraduate medical education. *Research and Development in Medical Education*, *11*, 22. https://doi.org/10.34172/rdme.2022.022

Rame, G., & Hoinbala, F. R. (2024). Item analysis of multiple-choice questions: An assessment of the English summative test. *Journal on Language, Education, Literature, and Culture*, *2*(1), 58–65. https://doi.org/10.33323/l.v2i1.45

Rashwan, N. I., Aref, S. R., Nayel, O. A., & Rizk, M. H. (2024). Postexamination item analysis of undergraduate pediatric multiple-choice questions exam: implications for developing a validated question Bank. *BMC Medical Education*, *24*(1), 168. https://doi.org/10.1186/s12909-024-05153-3

Rathnayake, B. R. M. A. S. M., & Senevirathne, W. (2024). The Relationship between Competency in Interpreting Data and Evidence Scientifically and The Academic Performance of Students in Biological and Physical Science Streams. *International Journal of Research and Scientific Innovation*, *XI*(II), 535–544. https://doi.org/10.51244/IJRSI.2024.1102043

Sarini, P., Widodo, W., Sutoyo, S., Suardana, I. N., & Info, S. (2024). Scientific Literacy Profile of Prospective Science Teacher Students. *IJORER : International Journal of Recent Educational Research*.

Schüttpelz-Brauns, K. (2022). Does assessment drive learning, or should learning for professional life drive assessment? *GMS Journal for Medical Education*, *39*.

Seema, P. V. (2024). Developing scientific literacy to promote 21st-century skills. *I-Manager's Journal on School Educational Technology*.

Srisomsak, V., Sitticharoon, C., Keadkraichaiwat, I., Meethes, S., & Inpaen, I. (2025). Detection of flawed multiple-choice questions in preclinical medical education using item difficulty and discrimination indices: a six-year analysis. *BMC Medical Education*, *26*(1), 92. https://doi.org/10.1186/s12909-025-08204-5

Steinbach, M., Eitemüller, C., Rodemer, M., & Walpuski, M. (2025). The influence of representations on task difficulty in organic chemistry: an exploration using a novel paired-items test instrument. *International Journal of Science Education*, *47*(14), 1941–1963. https://doi.org/10.1080/09500693.2024.2378218

Tani, E. W., & Nchia, L. N. (2025). An item difficulty analysis of the 2023 CEP and FSLC examinations in the Center Region of Cameroon: Implications for valid assessment practices in basic education. *International Journal of Assessment and Evaluation in Education*. https://doi.org/10.70382/mejaee.v9i8.051

Tirumalamba, D. K., Y.Sridevi, D. Rao, D. M. V., & Neelima, K. (2024). Teaching Socio-Scientific Issues Through Evidence-Based Thinking Practices. *Nanotechnology Perceptions*.

Wahyuni, S., Barid, S. S. A. Q., Gunawan, A. S., Rusdiana, D. U., & Sya'adah, H. (2026). Development of Integrated Case-Based Learning Science Modules on Coconut Shell Waste Management to Improve Science Literacy. *Jurnal Pendidikan MIPA*, *27*(2), 935–959. https://doi.org/10.23960/jpmipa.v27i2.pp935-959

Yanto, N., Sari, N. I., Negeri Makassar, U., & Lutra. (2025). Exploring Scientific Literacy in Science Classrooms: A Literature Study. *Venn: Journal of Sustainable Innovation on Education, Mathematics and Natural Sciences*.

Yudhistian, & Wulandari, T. S. H. (2025). Profile Analysis of Scientific Literacy Skills in Terms of Competencies, Knowledge, and Attitudes. *Jurnal Penelitian Pendidikan IPA*, *11*(6), 820–832. https://doi.org/10.29303/jppipa.v11i6.11580

Zetterqvist, A., & Bach, F. (2023). Epistemic knowledge – a vital part of scientific literacy? *International Journal of Science Education*, *45*(6), 484–501. https://doi.org/10.1080/09500693.2023.2166372

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

Zuleika, P., & Legiran. (2022). Cross-Sectional Study as Research Design in Medicine. *Archives of The Medicine and Case Reports*, *3*(2), 256–259. https://doi.org/10.37275/amcr.v3i2.193