

Mapping of scientific literacy instruments as a foundation for developing a transformative chemistry learning evaluation course model

Nahadi ^{1*} , Rika Sofiyatul Huda ¹ , Triannisa Rahmawati ¹ , Tri Lestari ¹ ,
Eki Nugraha ¹ , Salsabila Raniyah Mumtaz ² , Dini Dwi Apriani ¹ 

¹ Universitas Pendidikan Indonesia, Bandung, West Java, INDONESIA

² Universitas Padjadjaran, Bandung, West Java, INDONESIA

Received 31 May 2026 • Accepted 06 September 2026

Abstract

Scientific literacy assessment is a crucial aspect of science education, yet studies mapping the breadth of its instrument development remain limited. This study maps scientific literacy assessment instrument development, focusing on instrument types and formats, scientific domains, and research gaps. A scoping review was conducted through Scopus and ERIC databases, yielding 707 articles; 12 met the inclusion criteria and were analyzed using thematic synthesis. Scientific literacy is bounded here by a multidimensional framework of scientific competence, knowledge, and context, distinguishing it from narrower related constructs such as attitudes toward science or civic/digital scientific literacy. Results show that Likert-scale questionnaires and multiple-choice tests each dominate at 42% of studies. The general science (multi-dimensional) approach accounts for 67% of studies, while scenario-based and digital-based assessments remain rare. Integration of cognitive, affective, and social aspects into a single instrument is still limited, and psychometric quality varies across studies, highlighting the need for more integrative, authentic, and technology-based instruments aligned with 21st century learning demands.

Keywords: scientific literacy, assessment instruments, instrument development, scoping review

INTRODUCTION

Scientific literacy is recognized as an essential competency in 21st century education, playing a crucial role in equipping individuals to address complex global issues. Scientific literacy encompasses not only an understanding of scientific concepts but also the ability to apply that knowledge in responsible decision-making and in interpreting science-based issues in everyday life (Bybee, 1997; Organization for Economic Co-operation and Development [OECD], 2019). In this context, scientific literacy is understood as a multidimensional construct involving knowledge, skills, attitudes, and contexts of application (Zhang et al., 2023). OECD (2019) defines scientific literacy as an individual's capacity to engage reflectively with science-related issues, encompassing the competencies of explaining phenomena scientifically and evaluating empirical

evidence. This understanding has encouraged many countries to adopt scientific literacy as a primary goal of science education (Mardhiyyah et al., 2016; Rusilowati et al., 2016).

Because scientific literacy has been used in the literature to refer to a range of related but conceptually distinct constructs including scientific knowledge, scientific competence, scientific inquiry, scientific reasoning, nature of science (Lederman, 1992), scientific attitudes (Osborne et al., 2003), civic scientific literacy (Miller, 1998), digital scientific literacy and perceptions of scientific literacy this study explicitly bounds its scope before proceeding. In particular, awareness of or attitude toward scientific literacy reflects a dispositional or affective orientation (Osborne et al., 2003) and is not treated here as synonymous with the ability to perform scientific literacy. This review adopts the PISA multidimensional framework (OECD, 2019) as its

Contribution to the literature

- This study provides a comprehensive mapping of scientific literacy assessment instrument development, an aspect not addressed in prior reviews on this topic. Unlike previous reviews focused solely on theoretical frameworks, this study also maps instrument types, formats, and scientific domains, offering a broader analytical perspective.
- The findings identify critical gaps in the field including the limited adoption of authentic, digital-based, and integrative assessment approaches and propose concrete directions for future research.
- These contributions offer a practical roadmap for researchers and curriculum developers seeking to align scientific literacy assessments with 21st century competency demands.

guiding theoretical lens, under which scientific literacy is defined through three interrelated dimensions: scientific competence, scientific knowledge (which encompasses an understanding of the nature of science), and context. Civic scientific literacy, originally conceptualized by Miller (1998) as the level of scientific understanding needed to function as an informed citizen and more recently reconceptualized by Howell and Brossard (2021), is treated in this review as a context-specific application of this same underlying competence and knowledge, consistent with the PISA framework's contextual dimension, rather than as a separate construct; digital scientific literacy is treated analogously as its application to digitally mediated information. Scientific attitudes and perceptions of scientific literacy are treated as a distinct affective dimension (Osborne et al., 2003). Studies included in this review were therefore selected and classified based on whether the instrument was explicitly framed by its authors by measuring at least one of these dimensions of scientific literacy, rather than a related but distinct construct presented in isolation.

Although scientific literacy has been widely conceptualized, how best to develop it for all students remains a persistent challenge (Alonzo & Gotwals, 2012), and assessment plays a central role in addressing that challenge. High-quality instruments capture not only cognitive aspects but also scientific reasoning, attitudes, and evidence-based decision-making (Fives et al., 2014; Osborne et al., 2003), yet not all instrument-development research is grounded in a consistent and explicit theoretical framework, with implications for the validity and reliability of the resulting instruments (Istyadji & Sauqina, 2023). Approaches such as the Rasch model and confirmatory factor analysis (CFA) can provide important evidence of an instrument's internal structure and item characteristics, contributing to instrument validity and reliability (Bond & Fox, 2015), however, such statistical evidence does not by itself replace other necessary sources of validity evidence, and their application remains inconsistent across studies. Digital technology has further expanded the possibilities for scientific literacy assessment, enabling more adaptive, interactive, and real-time measurement of students' thinking processes (Shute & Ventura, 2013), yet the use

of digital assessment tools remains limited in the literature.

Reviewing instrument development studies is important for identifying the aspects of scientific literacy most frequently assessed, the assessment approaches employed, and the development processes used (Coppi et al., 2023; Fives et al., 2014). Istyadji and Sauqina (2023) conducted a systematic theoretical review focused on the conceptions and theoretical frameworks underlying scientific literacy instrument development but did not systematically map the types and formats of instruments produced, the science domains they targeted, or patterns of research gaps emerging across studies; these remaining gaps form the basis for this research.

Based on this background, this study aims to identify the types and formats of instruments used in scientific literacy assessment development, to map the science domains most frequently targeted and how the included studies conceptualized scientific literacy, and to identify research gaps and future directions in this field. The first two aims are addressed through two research questions (RQs):

RQ1. What types and formats of instruments are used in the development of scientific literacy assessment instruments?

RQ2. What science domains are most frequently focused on, and how is scientific literacy conceptualized, in the development of scientific literacy assessment instruments?

The third aim is addressed in the discussion, where research gaps and future directions are synthesized from the findings for **RQ1** and **RQ2**.

METHOD

Research Design

Scoping reviews are well suited to determining the scope and characteristics of a body of literature on a given topic, offering a broad overview of what has been studied without appraising the methodological quality of individual studies (Arksey & O'Malley, 2005). Given that this study aims to map the types, formats, and domains of scientific literacy assessment instruments and to identify emerging research gaps, a scoping review

design was adopted, following the methodological framework of Arksey and O'Malley (2005). This review was reported in accordance with the PRISMA 2020 guidelines (Page et al., 2021), adapted for the scoping-review format. The review was carried out through four sequential stages: identification, screening, eligibility, and inclusion, each designed to ensure that only articles relevant to the research focus were selected, particularly those addressing instrument types and formats, scientific literacy domains, and research gaps in instrument development.

Data Sources and Search Strategy

Literature searches were conducted through the Scopus and ERIC databases, selected for their extensive coverage of high-quality, peer-reviewed international publications and education-focused literature, respectively. Both databases were searched using the most recent version of the Publish or Perish software, applying the search string to the keyword field and filtering results by publication year (2016-2026). The final search was conducted in May 2026. Because Publish or Perish retrieval limits to 200 records per query, the Scopus search was conducted in successive two-year windows (2016-2018, 2018-2020, 2020-2022, 2022-2024, and 2024-2026) rather than as a single query, so that no eligible record was omitted due to the platform's per-query retrieval limit; results from each window were then merged. Duplicate records identified across the two databases were removed manually based on matching titles, authors, and publication years. In addition to the database search, the reference lists of the included studies were manually checked to identify further eligible records not captured by the search string.

The search string used in this study was as follows: ("scientific literacy" OR "science literacy") AND ("assessment" OR "instrument" OR "measurement" OR "scale" OR "test") AND ("development" OR "validation" OR "design").

This search strategy was designed to retrieve studies specifically related to the development and validation of scientific literacy assessment instruments. It should be noted that the search string did not include several terms commonly used in instrument-development literature such as "questionnaire," "inventory," "psychometric," "Rasch," "item response theory," "performance assessment," "rubric," "diagnostic test," and "scale development" that do not co-occur with "assessment," "instrument," "measurement," "scale," or "test" in some studies' titles, abstracts, or keywords. Consequently, this search may have systematically underrepresented studies that describe their instrument primarily through these alternative terms, and the resulting set of 12 studies should be read as illustrative of dominant approaches rather than an exhaustive census of the field.

Study Selection Process

The study selection process followed four sequential stages: identification, screening, eligibility, and inclusion. The screening and eligibility processes were conducted based on the predefined inclusion and exclusion criteria described above.

The inclusion criteria were as follows:

- (1) Empirical studies reporting the development and/or validation of a scientific literacy assessment instrument, operationally defined as a study that constructs a new instrument or substantially adapts an existing one, accompanied by at least one round of empirical validation (e.g., piloting, exploratory factor analysis (EFA)/CFA, Rasch analysis, or expert judgment), although the depth and completeness with which this validation process was reported varied across the included studies. This definition also encompasses instruments validated through qualitative or performance-based approaches (e.g., expert judgment, content validation, or classroom-based trialing), rather than exclusively through quantitative psychometric procedures, provided that some form of empirical trialing or expert validation was reported. Studies that merely administered a pre-existing, already-validated instrument without further development or validation work were excluded.
- (2) Articles published in peer-reviewed journals indexed in Scopus;
- (3) studies published between 2016 and 2026; and
- (4) studies related to science education, at any education level, including instruments targeting either students or teachers.

Instruments assessing scientific attitudes or perceptions were included where the authors explicitly framed them as measuring a dimension of scientific literacy (e.g., an affective dimension), consistent with the multidimensional PISA framework adopted in this review; general science-attitude scales not framed around scientific literacy were excluded. The exclusion criteria were as follows:

- (1) literature reviews, theoretical papers, or bibliometric studies;
- (2) studies that only administered a pre-existing instrument without reporting development or validation work;
- (3) studies focusing solely on learning outcomes without discussing assessment instruments;
- (4) attitude or perception scales not explicitly framed by the authors as measuring scientific literacy; and
- (5) conference proceedings, book chapters, and non-peer-reviewed publications.

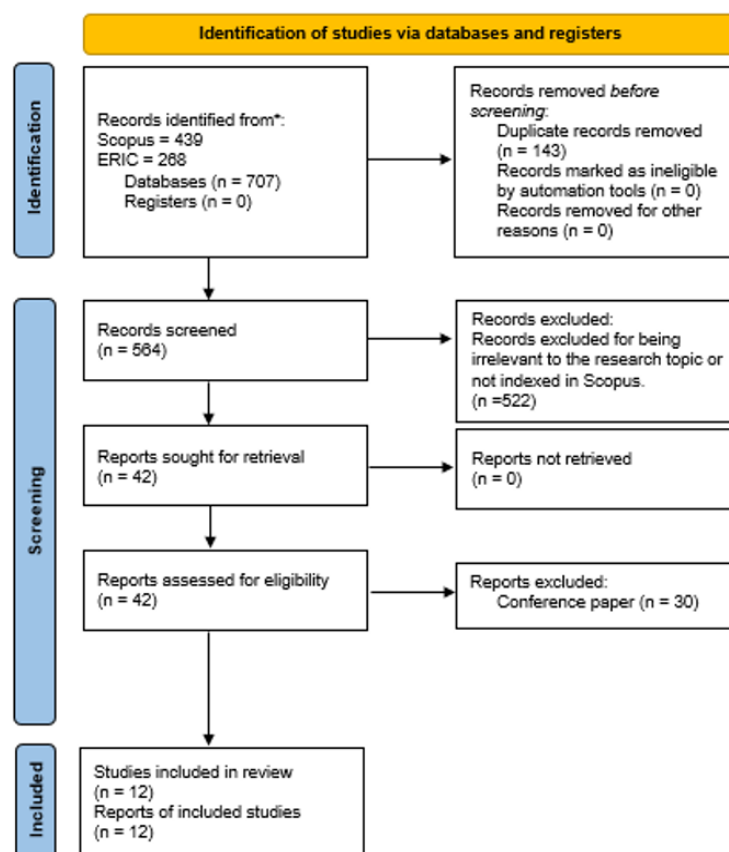


Figure 1. PRISMA diagram (the authors' own elaboration)

Although ERIC was searched to supplement bibliographic coverage, the Scopus-indexing requirement was retained as the primary inclusion criterion to ensure a consistent, verifiable minimum standard of peer review across all included studies. This criterion is nonetheless acknowledged as a limitation: journal indexing is not a direct indicator of a study's methodological quality, and this requirement may introduce indexing bias, exclude relevant studies published in non-Scopus-indexed regional or education-focused journals, and disproportionately affect studies from non-English-speaking regions.

Titles and abstracts of the retrieved articles were first screened to remove irrelevant studies. Subsequently, full-text analysis was conducted to assess eligibility based on the predefined criteria. Screening and eligibility assessment were conducted by the first author under the regular supervision of the second author, who was consulted on borderline or ambiguous cases. The final selection resulted in 12 articles that met all inclusion criteria and were included in the review.

The overall article selection process followed the PRISMA 2020 framework (Page et al., 2021), adapted for the scoping-review format, and is illustrated in **Figure 1**.

The initial search using the Scopus and ERIC databases identified a total of 707 articles based on the predefined search string (Scopus = 439; ERIC = 268). After removing 143 duplicate records, 564 articles

remained for the screening stage. During screening, titles and abstracts of the 564 records were reviewed based on relevance to the research topic and Scopus-indexing status. A total of 522 articles were excluded at this stage for being irrelevant to the research topic or not indexed in Scopus, and 42 articles were retained for the eligibility stage.

At the eligibility stage, a full-text assessment was conducted on the 42 selected articles based on the predefined inclusion and exclusion criteria. All 30 excluded articles were conference proceedings rather than peer-reviewed, Scopus-indexed journal articles, and were therefore excluded under exclusion criterion (5). Finally, 12 articles met all criteria and were included in the final analysis, as summarized in **Table 1**.

Data Extraction and Charting

A data-charting form was used to extract the following variables from each of the 12 included articles: author(s) and year, research method, instrument construct/context, content domain, response format, delivery mode, assessment purpose, education level of the target respondents, journal, and the study's main results and findings, as summarized in **Table 1**. Data extraction was performed by the first author. Where information in a source article was ambiguous or incomplete, the article was re-read in full, and, where necessary, supplementary materials accompanying the article were consulted to verify the relevant detail.

Table 1. Summary of analyzed articles

Study	Title	Method	Construct	Content domain	Context	Response format	Delivery mode	Assessment purpose	Journal	Key results
Mede et al. (2026)	Measuring science literacy in a digital world: Development and validation of a multi-dimensional survey scale	National quota surveys, Bayesian IRT & CFA	General scientific literacy (multi-dimensional competence)	General science	Civic, media, digital society	Questionnaire/ Likert scale (14 items)	Not specified	Summative	Science communication	The instrument captures three dimensions of scientific literacy relevant to a digital society civic, media, and cognitive literacy through a 14 item scale administered across national quota samples in Switzerland and the United States.
Suwono et al. (2022)	The development and validation of an instrument of prospective science teachers' perceptions of scientific literacy	EFA & CFA	Perception of science literacy	General science	Pre-service teacher education	Questionnaire/ Likert scale (41 items)	Not specified	Summative	Eurasia Journal of Mathematics, Science and Technology Education	The instrument measures prospective science teachers' perceptions of scientific literacy across eight dimensions, including their understanding of scientific literacy's role in science teaching.
Holincheck et al. (2022)	Assessing the development of digital scientific literacy with a computational evidence-based reasoning tool	Mixed methods & sInvestigator tool (evidence-evaluation competence)	Digital scientific literacy	General science	Digital/ online information context	Multiple-choice	Digital platform	Summative	Journal Educational Computing Research	The sInvestigator tool measures students' scientific literacy through a digital platform, focusing specifically on evaluating the credibility of online evidence.
Liu et al. (2024)	Development and validation of an instrument for measuring civic scientific literacy	R&D & validation	Civic scientific literacy competence	General science	Citizenship / civic context	50-item questionnaire, 5-point Likert scale	Not specified	Summative	Disciplinary and Interdisciplinary Science Education Research	The instrument measures scientific literacy specifically within citizenship contexts, extending measurement beyond cognitive understanding to civic application of scientific knowledge.
Coppi et al. (2023)	Developing a scientific literacy assessment instrument for Portuguese 3 rd cycle students	Content & internal structure validation	General scientific literacy (content knowledge)	General scientific literacy (physics, chemistry, biology)	General curriculum context	34-item true-false-don't know (ALCE)	Not specified	Summative	Education Sciences	The ALCE instrument aligns with the Portuguese science curriculum and assesses scientific literacy through content contextually relevant to 9 th grade students' everyday lives.
Zhang et al. (2023)	Development and validation of an instrument for assessing scientific literacy from junior to senior high school	Rasch model	General scientific literacy (content knowledge)	General scientific literacy (physics, chemistry, biology)	School curriculum context (elementary-senior high)	Multiple-choice, paper-based (PISA-like)	Paper	Summative	Disciplinary and Interdisciplinary Science Education Research	Using the Rasch model, the instrument classifies students' scientific literacy levels across elementary, junior, and senior high school, enabling comparison of literacy development across school levels.
Chang et al. (2024)	Introduction of Taiwanese literacy-oriented science curriculum and development of an aligned scientific literacy assessment	Curriculum alignment analysis	General scientific literacy (content knowledge)	General scientific literacy (physics, chemistry, biology)	Taiwanese curriculum context	Multiple-choice & case analysis (PISA-like)	Not specified	Summative	Eurasia Journal of Mathematics, Science and Technology Education	The instrument aligns with the Taiwanese literacy-oriented science curriculum, using multiple-choice and case-analysis items based on real-life contexts.
Ortega-Quevedo and Gil Puente (2020)	Formative assessment as a tool to highlight the development of skills in science and technology and critical thinking	Formative & qualitative assessment	Nature of science understanding	General science	Classroom/ formative learning context	Open-ended formative assessment (performance-based)	Classroom-based	Formative	Publications	The formative assessment approach monitors student competency development in critical thinking and scientific literacy by focusing on the learning process rather than a final outcome.
Benjamin et al. (2017)	Development and validation of scientific literacy scale for college preparedness in STEM with freshmen from diverse institutions	Psychometric analysis	General scientific literacy (STEM readiness)	General science	Higher-education transition context	Questionnaire/ Likert scale (SLSCP-STEM)	Not specified	Summative	International Journal Science and Mathematics Education	The SLSCP-STEM scale measures students' readiness for STEM learning at the point of transition into higher education.
Romine et al. (2017)	Assessment of scientific literacy: Development and validation of the quantitative assessment of socio-scientific reasoning (QuASSR)	Quantitative analysis (Rasch model)	Socio-scientific reasoning	General science	Societal decision-making context	Two-tiered multiple-choice (QuASSR)	Not specified	Summative	Journal Research in Science Teaching	The QuASSR instrument measures socio-scientific reasoning by separately assessing students' answer choices and the reasoning behind them, capturing critical thinking in science-related societal decision-making.

Table 1 (Continued).

Study	Title	Method	Construct	Content domain	Context	Response format	Delivery mode	Assessment purpose	Journal	Key results
Naganuma (2017)	An assessment of civic scientific literacy in Japan: Development of a more authentic assessment task and scoring rubric	Authentic rubric & task development	Civic scientific literacy competence	General science	Civic/societal (Japan) context	Scenario-based constructed-response (ACSEL) & rubric	Paper/task	Summative	International Journal Science Education, Part B	ACSEL measures scientific literacy as a civic competency through authentic, scenario-based tasks scored with a structured rubric, differing from conventional multiple-choice formats.
Çelik and Doğru (2024)	Development of an attitude scale towards science literacy for secondary school pupils	EFA & CFA	Scientific attitude (affective)	General science	Secondary-school context	16-item attitude scale	Not specified	Summative	Turkish Journal of Science Education	The scale measures secondary school students' attitudes toward scientific literacy as a distinct affective dimension, complementing cognitive-oriented instruments in the literature.

Data Analysis

Extracted data were analyzed descriptively (frequencies and percentages) to map instrument types, formats, and science domains (RQ1 and RQ2), and were further analyzed using thematic synthesis following Braun and Clarke (2006) to identify research gaps and future directions, discussed as a synthesis of the findings above. Coding combined a deductive structure derived from the RQs (instrument type/format, science domain, and research gap) with inductive, data-driven sub-codes generated from the included articles as they were read. Coding was carried out by the first author. Where a study's characteristics were relevant to more than one theme (e.g., an instrument spanning both cognitive and affective domains), it was reported under each relevant theme rather than being forced into a single exclusive category, to avoid loss of information. To avoid conflating distinct classification dimensions, findings are reported separately by construct/domain (RQ2), instrument type and delivery format (RQ1), and assessment purpose (discussed under research gaps), rather than within a single mixed category.

RESULTS

Studies on the development of scientific literacy assessment instruments over the past decade have demonstrated considerable variation in instrument type, format, and domain focus. The instruments developed reflect diverse approaches, ranging from conventional formats to more contextual and technology-based designs. The domains targeted also vary, encompassing both general and discipline-specific science literacy, and have begun to integrate cognitive and non-cognitive dimensions in alignment with the PISA-based scientific literacy framework comprising scientific competence, knowledge, and context (OECD, 2017). This variation suggests that scientific literacy assessment is a still-evolving field, underscoring the importance of systematically identifying mapping instrument types, formats, and development domains. The following discussion is organized around the two RQs guiding this review.

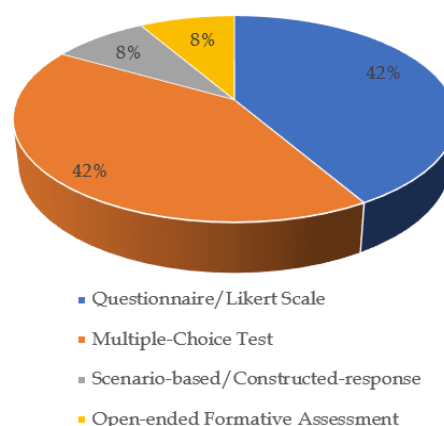


Figure 2. Distribution of instrument types and formats (the authors' own elaboration)

Instrument Types and Formats Used in Scientific Literacy Assessment

As shown in Figure 2, based on the extraction of the 12 analyzed articles the development of scientific literacy assessment instruments showed a relatively balanced distribution between two dominant formats: questionnaire-based instruments using a Likert scale and multiple-choice tests, each used in five studies (42%). Questionnaire-based instruments are generally employed to measure perceptions, attitudes, civic literacy, media literacy, cognitive literacy, and digital science literacy (Benjamin et al., 2017; Çelik & Doğru, 2024; Liu et al., 2024; Mede et al., 2026; Suwono et al., 2022). The prevalence of this approach is associated with its relative ease of development and statistical analysis, particularly in construct validation studies employing EFA and CFA (Bond & Fox, 2015). Questionnaire-based instruments also facilitate efficient large-scale data collection, making them the most common choice in scientific literacy scale development.

Multiple-choice instruments, by contrast, are used to more objectively measure conceptual understanding and scientific knowledge, primarily in the domains of biology, chemistry, and physics (Chang et al., 2024; Coppi et al., 2023; Zhang et al., 2023). Holincheck et al.

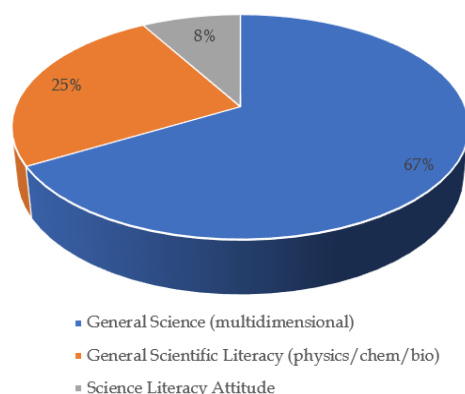


Figure 3. Distribution of scientific literacy domains (the authors' own elaboration)

(2022) employed a multiple-choice format delivered through a digital platform (the sInvestigator tool), demonstrating that conventional formats are beginning to be adapted to technology-based media even when the underlying question structure remains unchanged. Romine et al. (2017) developed the QuASSR instrument using a two-tiered multiple-choice format, in which each item consists of two levels: the first measures the answer choice and the second assesses the reasoning behind it. This approach is designed to measure socio-scientific reasoning more deeply than conventional multiple-choice, as it captures not only whether answers are correct but also the quality of the underlying reasoning.

Beyond these two dominant formats, two distinct approaches were each represented by only one study (8%). First, Naganuma (2017) developed the ACSEL using a scenario-based constructed-response format, in which participants were presented with a real-world socio-scientific scenario and asked to write open-ended responses assessed using a structured rubric with scores ranging from 0 to 3 an approach similar in design to PISA tasks. Second, Ortega-Quevedo and Gil Puente (2020) employed an open-ended formative assessment format consisting of reflective questions integrated into the learning process, analyzed qualitatively through observation categories without a standardized rubric. Although both approaches elicit open-ended written responses, they differ fundamentally in their level of standardization and purpose: the ACSEL was designed as a validated summative instrument, while Ortega-Quevedo and Gil Puente's (2020) approach is formative and contextual in nature. Both reflect efforts to capture more complex dimensions of scientific literacy including higher-order thinking, socially contextual decision-making, and the application of concepts to real-world situations that conventional formats are not well-suited to address.

Overall, these findings reveal a gap between scientific literacy as a multidimensional construct encompassing knowledge, competencies, and context (OECD, 2017) and current measurement practices, which remain

dominated by conventional formats that treat affective and cognitive aspects separately. While multiple-choice and questionnaire formats offer advantages in efficiency and ease of analysis, neither is fully capable of representing scientific literacy holistically. Although digital technology is beginning to drive the adoption of digital-based assessments such as interactive assessment systems and learning analytics (Ketelhut et al., 2023; Shute & Ventura, 2013), their use remains very limited in the literature.

Science Domains in Scientific Literacy Assessment Instrument Development

As shown in **Figure 3**, the analysis reveals that instrument development is dominated by the general science (multi-dimensional) approach, encompassing eight studies (67%). This group includes various subdomains such as civic/media/cognitive literacy (Mede et al., 2026) science literacy perception (Suwono et al., 2022), digital scientific literacy (Holincheck et al., 2022), civic scientific literacy (Liu et al., 2024; Naganuma, 2017), nature of science (Ortega-Quevedo & Gil Puente, 2020), STEM literacy (Benjamin et al., 2017), and socio-scientific reasoning (Romine et al., 2017). This approach reflects the view that scientific literacy is a multidimensional construct transcending specific disciplines and applicable across various life contexts including digital environments, citizenship, and socio-scientific issues (Fives et al., 2014; Howell & Brossard, 2021). According to Wenning (2006), scientific literacy encompasses intellectual skills such as problem-solving, reasoning, scientific inquiry, and critical thinking; conceptual knowledge of science, the nature of science, and the history of scientific development; and dispositions encompassing individuals' values, beliefs, attitudes, assumptions, and actions toward science.

The second group consists of discipline-based scientific literacy (biology, chemistry, and physics), represented by three studies (25%): Coppi et al. (2023), Zhang et al. (2023), and Chang et al. (2024). Instruments in this category are generally developed as multiple-choice tests that directly measure conceptual understanding and scientific competence, in alignment with the PISA framework (OECD, 2017). This approach tends to be more structured and standardized, facilitating comparisons across student groups and educational levels.

In terms of measurement foci, the discipline-based general scientific literacy domain emphasizes cognitive aspects in subjects such as biology, chemistry, and physics, measured primarily through multiple-choice tests. In contrast, the general science (multidimensional) domain integrates affective and contextual dimensions including civic literacy, media literacy, socio-scientific reasoning, nature of science, and digital science literacy measured through questionnaire-based instruments or open-ended formative assessments. These findings

indicate that instrument development remains predominantly oriented toward general approaches that treat cognitive and non-cognitive aspects separately and thus does not fully represent scientific literacy as an integrated construct (OECD, 2017).

Furthermore, only one study (8%) specifically targeted the affective dimension, through the development of a science literacy attitude scale (Çelik & Doğru, 2024). This finding suggests that the affective aspect of scientific literacy continues to receive limited attention relative to cognitive and contextual dimensions, despite being a recognized component of the holistic scientific literacy construct (Osborne et al., 2003). Overall, instrument development in the literature remains predominantly general in scope, while more contextual, reflective, and higher-order thinking-oriented instruments still require greater attention.

In terms of how scientific literacy itself was conceptualized, the included studies predominantly drew on multidimensional definitions consistent with the PISA framework of competence, knowledge, and context (OECD, 2019), though the specific sub-construct emphasized civic, digital, affective, or discipline-based varied considerably across studies.

DISCUSSION

Education Levels Targeted by the Instruments

The 12 included instruments also varied considerably in the education level of their target respondents. Higher-education populations were the most frequently targeted, represented by four studies: Suwono et al. (2022), targeting pre-service science teachers; Holincheck et al. (2022) and Benjamin et al. (2017), targeting college students; and Romine et al. (2017), targeting undergraduate students in environmental science courses. General adult or civic populations were targeted in three studies: Mede et al. (2026), surveying the general public in Switzerland and the United States; Liu et al. (2024), surveying adults aged 18 to 70; and Naganuma (2017), surveying Japanese citizens aged 20 and above. Secondary/junior high school students were targeted in two studies, Coppi et al. (2023) and Çelik and Doğru (2024), while elementary school students were targeted in two studies, by Chang et al. (2024) and Ortega-Quevedo and Gil Puente (2020). One study, Zhang et al. (2023), spanned multiple levels simultaneously, from elementary through senior high school.

This distribution suggests that instrument development has concentrated primarily on higher-education and general-population contexts, where large-scale surveys and psychometric validation are logistically easier to conduct, while dedicated instruments for elementary-level students remain comparatively rare, despite scientific literacy being an

explicit goal of science curricula from the earliest grades (OECD, 2019). Moreover, only Benjamin et al. (2017) developed an instrument specifically to assess readiness at the transition into higher education, indicating a gap in instruments designed to support students during critical educational transitions, when monitoring scientific literacy development could inform targeted pedagogical intervention. The concentration of instruments at the pre-service teacher level (Suwono et al., 2022) rather than the in-service level also points to a further gap: comparatively little attention has been given to assessing scientific literacy among practicing teachers, who directly shape how scientific literacy is taught and assessed in classrooms.

Research Gaps and Future Directions

The analysis reveals several persistent gaps in the development of scientific literacy assessment instruments, indicating that the field has not yet reached comprehensive maturity. The findings from **RQ1** and **RQ2** show that although multidimensional approaches are beginning to dominate (67%), instrument development overall has not yet succeeded in representing scientific literacy as a fully integrated construct.

First, instrument development remains dominated by conventional formats Likert-scale questionnaires and multiple-choice tests, each used in 42% of the analyzed studies. The use of performance-based assessment, authentic assessment, and digital-based instruments remains very limited (de Moraes & Rodrigues, 2021; Fives et al., 2014). Of the 12 articles analyzed, only Naganuma (2017) employed a scenario-based constructed-response approach with a structured rubric, and only Holincheck et al. (2022) fully integrated digital platforms into the assessment process. This indicates that scenario-based and technology-based assessments have not yet become mainstream approaches in scientific literacy instrument development (Wiggins, 1999).

Second, although the general science (multidimensional) domain dominates with 67% of studies, most development efforts are still conducted separately by subdomain such as civic literacy, digital scientific literacy, socio-scientific reasoning, and nature of science without integrating these dimensions into a single, coherent instrument. Integration of cognitive, affective, and social aspects within a single instrument remains very limited (OECD, 2019). Additionally, affective aspects such as attitudes toward scientific literacy are represented by only one study (8%) (Çelik & Doğru, 2024), and metacognitive dimensions such as teacher perceptions of scientific literacy by only one study (Suwono et al., 2022), indicating that these dimensions continue to receive insufficient attention (Fives et al., 2014).

Third, the integration of digital technology into scientific literacy assessments has not yet developed optimally. Although Holincheck et al. (2022) developed a computational tool-based instrument (sInvestigator) and Mede et al. (2026) measured scientific literacy in the context of a digital society, skills related to digital literacy such as evaluating the credibility of online information sources are rarely addressed comprehensively in existing instruments (Ketelhut et al., 2023; Shute & Ventura, 2013). Yet these skills are increasingly critical in 21st century learning environments saturated with digital information.

Fourth, the included instruments were validated using diverse approaches. Several studies employed quantitative psychometric methods such as EFA and CFA (Çelik & Doğru, 2024; Mede et al., 2026; Suwono et al., 2022) and the Rasch model (Romine et al., 2017; Zhang et al., 2023), while Ortega-Quevedo and Gil Puente (2020) employed qualitative analysis without a standardized rubric. This diversity reflects differing purposes and contexts of instrument development, from large-scale quantitative measurement to classroom-based formative assessment, rather than a single agreed upon validation standard for scientific literacy instruments.

Fifth, instrument development is not yet well aligned with the principles of continuous, formative assessment (assessment for learning). This is relevant to scientific literacy because scientific literacy develops gradually as students engage in learning, rather than being demonstrated fully at a single point in time; continuous assessment allows this development to be tracked throughout the learning process, whereas a single summative measurement only captures the end result (Black & Wiliam, 2009). Most of the instruments analyzed are oriented toward assessment of learning, administered as a single, summative measurement, while an assessment for learning approach that continuously monitors students' developing scientific literacy across the learning process was found in only one study, by Ortega-Quevedo and Gil Puente (2020), through formative assessment.

Sixth, the development of instruments that measure higher-order thinking skills such as science-based decision-making, argumentation, and evidence evaluation has not been conducted comprehensively. Although some studies, such as Romine et al. (2017) through the QuASSR and Naganuma (2017) through the ACSEL, have attempted to address these dimensions, they remain rare. Yet these skills are central to scientific literacy within the 21st century skills framework and international assessments such as PISA (OECD, 2019). The development of authentic, problem-based assessments with contextual rubrics therefore remains an urgent priority for future research.

Based on these identified gaps, future research should focus on developing more integrative, contextual, and technology-based scientific literacy assessment instruments. For researchers, these findings provide a clear agenda for areas requiring attention particularly the development of instruments capable of measuring scientific literacy holistically by combining cognitive, affective, and social dimensions within a single, cohesive instrument. For educators and curriculum developers, the results of this study underscore the importance of aligning assessment instruments with 21st century competency demands, including the integration of digital literacy, the strengthening of assessment for learning practices, and the development of authentic assessments grounded in real-world problems.

CONCLUSION

This review maps the current landscape of scientific literacy assessment instrument development based on 12 identified studies, revealing considerable variation in instrument type, format, and measured domain. Instrument development is dominated by two main formats: Likert-scale questionnaires and multiple-choice tests, each used in 42% of the analyzed studies. Questionnaires are generally employed to measure perceptions, attitudes, and multidimensional aspects of scientific literacy, while multiple-choice tests are more commonly used to objectively measure conceptual understanding. By contrast, more contextual and authentic formats such as scenario-based constructed-response and open-ended formative assessments remain very limited, each found in only one study (8%). Similarly, digital technology-based instruments have not yet emerged as a primary approach in scientific literacy instrument development.

In terms of domain, instrument development is dominated by the general science (multidimensional) approach, representing 67% of the analyzed studies, with diverse subdomains including civic scientific literacy, digital scientific literacy, socio-scientific reasoning, nature of science, STEM literacy, science literacy perception, and civic/media/cognitive literacy. Discipline-based scientific literacy (biology, chemistry, and physics) accounts for 25% of studies, while instruments specifically targeting the affective dimension remain very limited at only 8%. Although the multidimensional approach is beginning to dominate, the integration of cognitive, affective, and social dimensions of scientific literacy into a single, comprehensive instrument has not yet been widely implemented.

The development of scientific literacy assessment instruments continues to exhibit several significant gaps, including the dominance of conventional formats, limited digital technology integration, uneven

psychometric quality, and a scarcity of instruments oriented toward assessment for learning and higher-order thinking skills. Future research should therefore prioritize the development of more integrative, authentic, and technology-based instruments capable of measuring scientific literacy comprehensively combining cognitive, affective, and social dimensions in a balanced manner in alignment with the demands of 21st century education. Such instruments could be made more digitally based through interactive scenario simulations, automated evaluation of online-information credibility, or embedded assessment within digital learning platforms, rather than relying solely on digitized versions of conventional item formats.

Limitations

This study has several limitations. The literature search was restricted to two databases (Scopus and ERIC). The search string was deliberately restricted to terms directly denoting assessment, instrument, measurement, scale, and test combined with development, validation, or design, in order to capture studies that explicitly frame their contribution as instrument development or validation work. Related terms such as questionnaire, psychometric, Rasch, item response theory, rubric, and diagnostic test were intentionally not added, as many of these terms return a large proportion of studies that use such instruments or methods without developing or validating a new one, which would have substantially increased screening burden without a proportionate increase in eligible records; we acknowledge this as a boundary of the present review's search strategy rather than an exhaustive search of the field. The Scopus-indexing requirement may also have introduced indexing bias, excluding relevant studies published in non-Scopus-indexed regional or education-focused journals. The relatively small number of final articles ($n = 12$) and the restriction to English-language publications from 2016 to 2026 further limit the generalizability of the findings. Consequently, the percentages and patterns reported throughout this review describe this identified set of studies rather than a definitive or exhaustive account of the field.

Screening, eligibility assessment, data extraction, and thematic coding were conducted by a single reviewer rather than independently by two or more reviewers, which may introduce subjectivity; inter-rater reliability indices could not be computed as a result. As a scoping review, this study also did not independently assess the psychometric quality of the included instruments; validity and reliability claims reported here reflect what was reported by the original authors. These limitations should be considered when interpreting the results and may serve as directions for future studies to involve multiple independent reviewers and broaden the search scope, languages, and databases.

Author contributions: **N:** conceptualization, supervision, validation, writing – review & editing; **RSH:** conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing – original draft, writing – review & editing; **TR:** investigation, project administration, writing – review & editing; **TL & DDA:** investigation, writing – review & editing; **EN & SRM:** investigation, writing – review & editing. All authors agreed with the results and conclusions.

Funding: This study was funded by the Universitas Pendidikan Indonesia through the research scheme/program of Riset Penugasan Kelembagaan with contract no. 57/UN40.F7/ADD-PT.01.01/2026).

Ethical statement: This study is a systematic literature review and did not involve direct participation of human subjects. Therefore, formal ethical approval was not required.

AI statement: During the preparation of this work, the authors used Claude AI to assist with improving language, clarity, and readability, as well as to support the organization and structuring of ideas throughout the manuscript. The authors reviewed, verified, and edited all AI-assisted content, and took full responsibility for the accuracy, integrity, and originality of the published work.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Alonzo, A. C., & Gotwals, A. W. (2012). *Learning progressions in science: Current challenges and future directions*. Sense Publishers. <https://doi.org/10.1007/978-94-6091-824-7>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Benjamin, T. E., Marks, B., Demetrikopoulos, M. K., Rose, J., Pollard, E., Thomas, A., & Muldrow, L. L. (2017). Development and validation of scientific literacy scale for college preparedness in STEM with freshmen from diverse institutions. *International Journal of Science and Mathematics Education*, 15(4), 607-623. <https://doi.org/10.1007/s10763-015-9710-x>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315814698>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Heinemann.

- Chang, T.-C., Lyu, Y.-M., Wu, H.-C., & Min, K.-W. (2024). Introduction of Taiwanese literacy-oriented science curriculum and development of an aligned scientific literacy assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(1), Article em2380. <https://doi.org/10.29333/ejmste/13930>
- Coppi, M., Fialho, I., & Cid, M. (2023). Developing a scientific literacy assessment instrument for Portuguese 3rd cycle students. *Education Sciences*, 13(9), Article 941. <https://doi.org/10.3390/educsci13090941>
- Çelik, M., & Doğru, M. (2024). Development of an attitude scale towards science literacy for secondary school pupils. *Journal of Turkish Science Education*, 21(2), 304-323. <https://doi.org/10.36681/tused.2024.017>
- De Moraes, J. V., & Rodrigues, P. B. (2021). *Scientific literacy in geography: The use of indicators for the assessment of learning*. Springer. https://doi.org/10.1007/978-3-030-79847-5_12
- Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M. (2014). Developing a measure of scientific literacy for middle school students. *Science Education*, 98(4), 549-580. <https://doi.org/10.1002/sce.21115>
- Holincek, N., Galanti, T. M., & Trefil, J. (2022). Assessing the development of digital scientific literacy with a computational evidence-based reasoning tool. *Journal of Educational Computing Research*, 60(7), 1796-1817. <https://doi.org/10.1177/07356331221081484>
- Howell, E. L., & Brossard, D. (2021). (Mis)informed about what? What it means to be a science-literate citizen in a digital world. *PNAS*, 118(15), 1-8. <https://doi.org/10.1073/pnas.1912436117>
- Istyadji, M., & Sauqina. (2023). Conception of scientific literacy in the development of scientific literacy assessment tools: A systematic theoretical review. *Journal of Turkish Science Education*, 20(2), 281-308. <https://doi.org/10.36681/tused.2023.016>
- Ketelhut, D. J., Nelson, B. C., Clarke, J., & Dede, C. (2023). Inq-ITS: Creating rigorous assessment and real-time support of science learning. *Journal of the Learning Sciences*, 29(3), 36-40. <https://doi.org/10.1145/3589647>
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359. <https://doi.org/10.1002/tea.3660290404>
- Liu, Y., Wang, J., Zhang, Z., Wang, J., Luo, T., Lin, S., Li, J., & Xu, S. (2024). Development and validation of an instrument for measuring civic scientific literacy. *Disciplinary and Interdisciplinary Science Education Research*, 6, Article 6. <https://doi.org/10.1186/s43031-023-00092-3>
- Mardhiyyah, L. A., Rusilowati, A., & Linuwih, S. (2016). Pengembangan instrumen asesmen literasi sains tema energi [Development of scientific literacy assessment instruments on the theme of energy]. *Journal of Primary Education*, 5(2), 147-154. <https://journal.unnes.ac.id/sju/jpe/article/view/12905/7042>
- Mede, N. G., Howell, E. L., Schäfer, M. S., Metag, J., Beets, B., & Brossard, D. (2026). Measuring science literacy in a digital world: Development and validation of a multi-dimensional survey scale. *Science Communication*, 48(1), 93-127. <https://doi.org/10.1177/10755470251317379>
- Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, 7(3), 203-223. <https://doi.org/10.1088/0963-6625/7/3/001>
- Naganuma, S. (2017). An assessment of civic scientific literacy in Japan: Development of a more authentic assessment task and scoring rubric. *International Journal of Science Education, Part B: Communication and Public Engagement*, 7(4), 301-322. <https://doi.org/10.1080/21548455.2017.1323131>
- OECD. (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. OECD Publishing. <https://doi.org/10.1787/9789264281820-en>
- OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Ortega-Quevedo, V., & Gil Puente, C. (2020). La evaluación formativa como elemento para visibilizar el desarrollo de competencias en ciencia y tecnología y pensamiento crítico [Formative assessment as a tool to highlight the development of skills in science and technology and critical thinking]. *PUBLICACIONES*, 50(1), 275-291. <https://doi.org/10.30827/publicaciones.v50i1.15977>
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049-1079. <https://doi.org/10.1080/0950069032000032199>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Romine, W. L., Sadler, T. D., & Kinslow, A. T. (2017). Assessment of scientific literacy: Development and validation of the quantitative assessment of socio-scientific reasoning (QuASSR). *Journal of Research in Science Teaching*, 54(2), 274-295. <https://doi.org/10.1002/tea.21368>
- Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an instrument of scientific literacy assessment on the cycle theme. *International Journal of Environmental and Science Education*, 11(12), 5718-5727. <https://eric.ed.gov/?id=EJ1115684>
- Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press. <https://doi.org/10.7551/mitpress/9589.001.0001>
- Suwono, H., Maulidia, L., Saefi, M., Kusairi, S., & Yuenyong, C. (2022). The development and validation of an instrument of prospective science teachers' perceptions of scientific literacy. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(1), Article em2068. <https://doi.org/10.29333/EJMSTE/11505>
- Wenning, C. J. (2006). Assessing nature-of-science literacy as one component of scientific literacy. *Journal of Physics Teacher Education Online*, 3(4), 3-10.
- Wiggins, G. (1999). Educative assessment: Designing assessments to inform and improve student performance. *Choice Reviews Online*, 36(5). <https://doi.org/10.5860/choice.36-2887>
- 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, Article 21. <https://doi.org/10.1186/s43031-023-00093-2>

<https://www.ejmste.com>