Al-Ishlah: Jurnal Pendidikan
Vol.17, 3 (September, 2025), pp. 3787-3802
ISSN: 2087-9490 EISSN: 2597-940X, DOI: 10.35445/alishlah.v17i3.6198

Trend in Science Education Assessment Instruments: A Systematic Literature Review (2014-2024)

Anisa Ayu Solikah[1], Sulistyo Saputro*[2], Sri Yamtinah[3]

[1] Universitas Sebelas Maret, Surakarta, Indonesia; anisaayusolikah2023@student.uns.ac.id
[2] Universitas Sebelas Maret, Surakarta, Indonesia; sulistyo_s@staff.uns.ac.id
[3] Universitas Sebelas Maret, Surakarta, Indonesia; jengtina@staff.uns.ac.id

ARTICLE INFO

Keywords:

systematic literature review;
assessment instruments;
science education;
educational trends (2014-2024)

Article history:

Received 2024-11-02
Revised 2025-01-20
Accepted 2025-08-25

ABSTRACT

Assessment in education plays a critical role in evaluating not only students' cognitive understanding but also broader 21st-century skills such as critical thinking, collaboration, and communication. However, traditional assessments often prioritize conceptual knowledge over these essential skills. This study investigates current trends in assessment within science education to identify gaps and potential improvements. A Systematic Literature Review (SLR) was conducted using articles sourced from the Google Scholar database between 2014 and 2024. The search employed keywords such as "assessment learning" and "science learning." A total of 992 articles were initially collected, which were then filtered based on relevance and quality, resulting in 76 eligible articles. Data analysis was supported by Publish or Perish software and visualized using VOSviewer. The VOSviewer analysis produced network, overlay, and density visualizations that revealed a predominant focus on conceptual understanding in science education assessments. This focus likely stems from the relative ease of standardizing and measuring conceptual knowledge compared to more complex 21st-century skills. Other dimensions, such as creativity, collaboration, and problem-solving, were significantly underrepresented in the literature. The findings highlight a need to diversify assessment strategies in science education to better reflect 21st-century learning goals. This review provides a foundation for future research aimed at developing comprehensive assessment instruments that go beyond cognitive evaluation and support holistic student development.

Corresponding Author:
Sulistyo Saputro
Universitas Sebelas Maret, Surakarta, Indonesia; sulistyo_s@staff.uns.ac.id

1. INTRODUCTION

Assessment is a systematic process used to evaluate individuals' and groups' knowledge, achievements, and competencies based on defined criteria and supporting evidence (Ministry of Education, 2011). In educational contexts, assessment is essential for understanding what students know and can do—a principle captured in the concept of "Knowing What Students Know" (National Research Council, 2001). Effective assessment is not merely an endpoint evaluation but an ongoing

process that informs instruction and supports learning outcomes. It helps educators make informed decisions regarding teaching strategies, student progress, and curriculum effectiveness (Amelia, Mutmainah, Indriyani, & Carmenette, 2023). When well-aligned with educational goals and context, assessments serve as vital tools for guiding both teaching and learning (Almossa & Alzahrani, 2022).

The quality of assessment practices directly influences the success of educational implementation. Poorly designed assessments can hinder lesson effectiveness, misrepresent individual achievement, and undermine overall learning outcomes (Songer & Ruiz-Primo, 2012). One effective approach is "assessment for learning," a refined form of formative assessment that fosters meaningful interaction between teachers and students, promoting continuous feedback and learning improvement (Kaptan & Timurlenk, 2012). Assessments also function as validation tools to ensure that learning objectives are being met (Amelia et al., 2023), particularly those aligned with 21st-century skills such as critical thinking, collaboration, and digital literacy.

Despite their importance, many teachers face significant challenges in designing assessments that reflect both curriculum standards and real-world demands. Often constrained by time, training, and institutional requirements, educators tend to rely on standardized or previously published assessments, many of which emphasize conceptual understanding through essay and multiple-choice formats (Bernard, Abrami, Borokhovski, Wade, Tamim, Surkes, & Bethel, 2008; Desriani, Takunas, & Nadira, 2021). While easier to administer and standardize, these forms of assessment primarily target lower-order thinking skills and do not adequately capture the complexity of 21st-century learning outcomes (Sumintono & Widhiarso, 2015).

In response to rapid societal and technological changes, 21st-century education must go beyond the simple transmission of knowledge. It must also develop a broad set of cognitive and non-cognitive skills, including critical thinking, problem-solving, creativity, communication, collaboration, and digital fluency (Anwar & Uman, 2020; Santos-Meneses & Drugova, 2023). These shifts are driven by global transformations in communication, information, transportation, and the economy that demand adaptive, lifelong learners (Yamin, 2019). Schools, therefore, have a critical responsibility to prepare students to navigate these dynamic conditions by cultivating emotional intelligence, resilience, and the ability to manage complexity (Nieto, 2017).

The importance of 21st-century skills is further emphasized by global educational frameworks. UNESCO's four pillars of learning—learning to know, learning to do, learning to be, and learning to live together—form the foundation for lifelong learning and global citizenship (Junedi, Mahuda, & Kusuma, 2020). These principles aim to foster well-rounded individuals who can contribute meaningfully to society. In alignment with these goals, Indonesia's Merdeka Curriculum promotes learner independence, critical and creative thinking, as well as collaboration and communication skills as central educational objectives (Sari, 2024). Consequently, teachers must not only teach subject-specific content but also embed 21st-century competencies into the learning process—especially in subjects like science.

Science is a field inherently concerned with both observable phenomena (facts) and the theories that explain them. According to Gould (1997), science seeks to understand "what is" and "why it is" through systematic inquiry. Its foundational characteristics—rationality, objectivity, and accumulation of knowledge—demand that students engage in logical reasoning, evidence-based thinking, and iterative exploration (Dettweiler, 2019). Ideally, science education should enable students to apply scientific thinking to real-world problems, helping them construct meaningful knowledge and develop analytical skills (Roorda, 2022).

However, the reality of science education often falls short of this ideal. Many students find science learning abstract and disconnected from everyday experiences. In disciplines such as chemistry and biology, students are frequently required to memorize information without truly understanding its application, while those studying physics and mathematics often struggle to see the relevance of abstract formulas to real-world scenarios (Kwok, 2018). These shortcomings extend to the

assessment process, where traditional evaluations often fail to capture deeper learning or higher-order thinking skills.

Given the urgent need to reform science education in line with 21st-century demands, assessment must evolve to reflect a broader range of learning outcomes. Current research reveals a gap in the literature concerning the development and use of science assessment instruments that measure both conceptual understanding and essential 21st-century skills. Most existing assessments remain narrowly focused on content mastery, with limited attention to skills like creativity, collaboration, or scientific reasoning.

Given the critical role of assessment in supporting 21st-century learning, a comprehensive understanding of the trends and development of assessment instruments in science education over the past decade is essential. This study aims to address that gap by conducting a Systematic Literature Review (SLR) of publications related to assessment in science education from 2014 to 2024. The review utilizes VOSviewer, a bibliometric analysis tool used to map co-citation networks, bibliographic coupling, and keyword co-occurrence to uncover research trends and thematic clusters (Van Eck & Waltman, 2021). The primary objective of this study is to identify how assessment practices in science education have been applied to measure variables relevant to 21st-century competencies and how these practices have evolved over time. The findings are expected to provide a solid foundation for future research and the development of more comprehensive and integrated assessment instruments. Ultimately, this study contributes to advancing science education by aligning assessment strategies with the broader goals of modern learning and equipping students to meet future societal and professional challenges.

2. METHODS

2.1 Research Method

This study employed a Systematic Literature Review (SLR) as the primary research method. A Systematic Literature Review is a structured approach to identifying, analyzing, and synthesizing research evidence on a specific topic or research question, allowing for a transparent, replicable, and unbiased process (Lame, 2019). This methodology incorporates a qualitative approach, where literature is systematically reviewed to compile descriptive findings relevant to the topic under investigation (Ahn & Kang, 2018). The goal is to understand research developments and trends in science education assessment over the last decade.

2.2 Data Collection Procedure

The data collection process involved a comprehensive literature search using the Publish or Perish (PoP) software, which extracts metadata from academic databases such as Google Scholar. This tool was selected for its effectiveness in gathering a large volume of academic literature based on specified keywords. The search was conducted using the keywords: “assessment learning” and “science learning”, covering the publication period from 2014 to 2024.

This initial search yielded a total of 992 articles. These articles were then subjected to a filtering process based on several inclusion criteria:

- Articles published in English,
- Peer-reviewed journal articles,
- Relevance to the fields of science education and educational assessment,
- Appropriate document types (e.g., articles, conference papers),
- Proper publication timeline (2014–2024).

After applying these criteria, 76 articles remained and were considered eligible for further analysis, as summarized in Table 1. Each article was carefully reviewed to assess its relevance to the topic and to identify specific themes, methods, and findings related to assessment in science education.

Of the 76 articles, 20 were found to directly focus on science education assessment instruments. These articles provided in-depth information for answering the research questions and served as the core data set for further analysis using VOSviewer.

2.3 Data Analysis Using VOSviewer

The final step involved analyzing the bibliometric data of the selected articles using VOSviewer, a software tool designed for constructing and visualizing bibliometric networks (Van Eck & Waltman, 2021). VOSviewer enables researchers to explore trends and relationships within large bodies of scientific literature through keyword co-occurrence, co-citation analysis, bibliographic coupling, and network visualization.

This tool was used to map patterns in the research on science education assessments over the 2014–2024 period, helping to identify dominant themes, frequent terminologies, and the interconnections among studies. The output includes network, overlay, and density visualizations, which provide insight into how the field of science assessment has developed and where further research is needed.

Table 1. Article criteria

Criteria	Inclusion	Exclusion
Research	In science education -Focusing on science education assessment	-outside science education -not focusing on learning assessment
Period	January 2014-January 2024	Before January 2014
Document Type	articles in the form of journals and proceeding about assessment in science education.	-Book Chapter -Book -Literature review
Language	English	Non-English
Publication Stage	Final	In Press

The research process began by identifying relevant keywords, specifically “assessment learning” and “science learning,” which were used to search for articles via the Google Scholar database and the Publish or Perish (PoP) software. The initial stage involved mining the literature by selecting and reading articles that aligned with the research objectives, with a focus on publications from 2014 to 2024. In the subsequent phase, article selection was refined using predefined inclusion criteria, as outlined in Table 1. The overall procedure of the systematic literature review, including data collection and screening, is visually summarized in Figure 1.

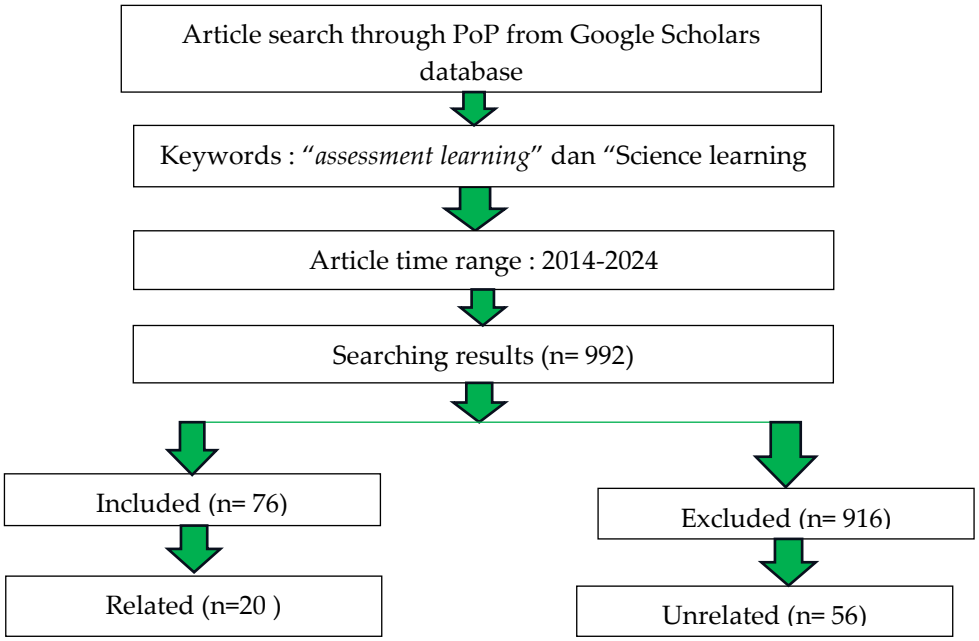

Figure 1. The procedure of systematic literature review analysis

The results of the literature search before and after the filtering process were compared to evaluate the refinement of the selection. Table 2 presents a comparison of the bibliometric data, showing an initial total of 992 articles, which was narrowed down to 20 highly relevant articles. The article selection process was conducted using Publish or Perish (PoP), followed by downloading and thoroughly reviewing the full texts to identify those that aligned most closely with the research objectives. The final set of 20 articles specifically focused on science learning assessment, particularly in the development of assessment instruments. These selected articles were then exported in RIS format and analyzed using VOSviewer to visualize bibliometric patterns and research trends.

Table 2. Ratio Metrics

Citation metrics	Initial research	Final Data Mining
Publication years	2012-2014	2014-2023
Citation years	12 (2012-2024)	10 (2014-2024)
Papers	992	20
Citations	36018	599
Cites/year	3001.50	59.9
Cites/paper	36.31	29.95
Cites/author	20072.86	202.85
Papers/author	503.37	7.90
Authors/paper	2.63	2.95
h-index	86	7
g-index	170	20
hI, norm	62	5
hI, annual	5.17	0.50
hA-index	32	5

3. FINDINGS AND DISCUSSION

In this research, researchers focused on trends and research assessment studies in science education from 2014 to 2024, supported by Vos viewer. Vos viewer results generally show that in

science learning, the assessments that are often carried out are to assess students' cognitive and understanding. Apart from that, the form of assessment that is most often carried out is in the form of multiple-choice questions. The discussion is as follows:

3.1. Trend assessment study in science education from 2014 to 2024

The analysis using VOSviewer generated three types of visualizations: network visualization, overlay visualization, and density visualization, as presented in Figures 2 to 4. From this analysis, a total of 54 items were identified and grouped into 5 clusters, forming 792 links with a total link strength of 1,186. These 54 items were derived from 20 selected articles containing the keywords *"assessment learning"* and *"science learning."* The items represent 54 interconnected terms that frequently co-occur in the selected literature. Each item is grouped into a specific cluster, distinguished by different colors, as detailed in Table 3.

Table 3. Items in every cluster

No.	Cluster (colour)	Element
1.	First cluster (red)	Ability, activity, classroom, concept, development, education, fact, information, learning process, learning tool, outcome, percentage, practicality, science learning tool, science process skill, syllabus, technology, validation, virtual laboratory
2.	Second cluster (green)	Achievement, effectiveness, elementary school, experiment, experimental group. Knowledge, learning achievement, motivation, peer assessment, perception, project, self-efficacy, study
3.	Third cluster (blue)	Assessment, chemistry, formative assessment, impact, interview, learning approach, questionnaire, student performance, test, understanding
4.	Fourth cluster (yellow)	high school student, implementation, instrument, portfolio assessment, reliability, science, science learning, validity
5.	Fifth cluster (purple)	Investigation, observation. Post-test, pretest

The article is analyzed based on Network Visualization, Density visualization, and Overlay Visualization. Network visualization contains items represented by labels and networks, where the size of the labels and item circles is determined by the number of items. The overlay visualization represents the year the research related to the item was carried out, which is aimed at colour changes, where the newer the research, the brighter (yellow) the colour of an item group. Meanwhile, density visualization represents the density of items, the greater the number of items, the closer the colour is to red (Solikah, Saputro, Yamtinah, & Masykuri, 2024).

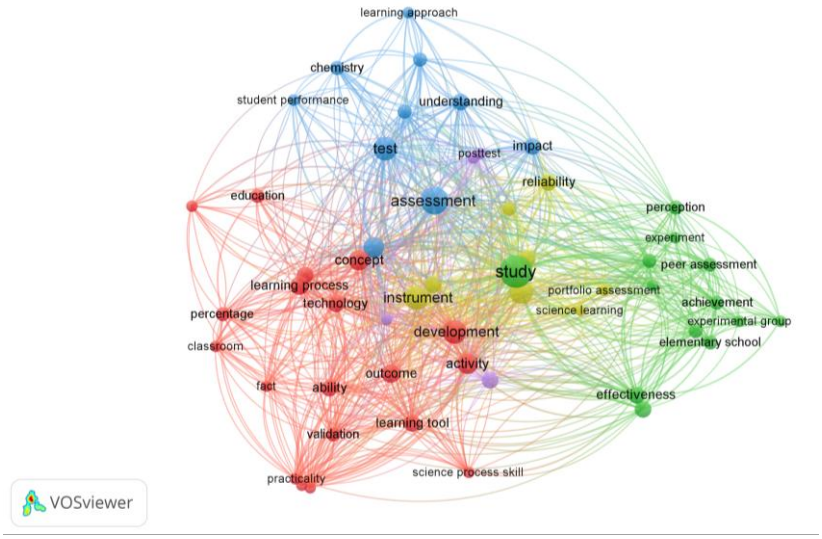

Figure 2. Network Visualization result

Figure 2 illustrates the correlation between keywords that remain highly relevant to the field of study. The term "assessment" appears as a central node in the network, strongly linked to keywords such as "science process skills," "understanding," "learning process," "portfolio assessment," "experiment," "ability," "outcome," "reliability," "pre-test," "post-test," "validation," "concept," "technology," and "impact." This centrality indicates that *"assessment"* is a pivotal concept within the literature, particularly in relation to cognitive domains such as conceptual understanding and science process skills.

The visualization confirms that research in science education assessment continues to emphasize cognitive outcomes, aligning with previous findings by Sumintono and Widhiarso (2015), who argued that the cognitive dimension is more frequently assessed due to its measurability and ease of standardization. In contrast, assessments targeting 21st-century skills—such as critical thinking, creativity, communication, collaboration, and global awareness—remain underrepresented in the literature. This gap highlights the importance of bibliometric tools like VOSviewer, which can uncover emerging themes and research blind spots within the existing body of work.

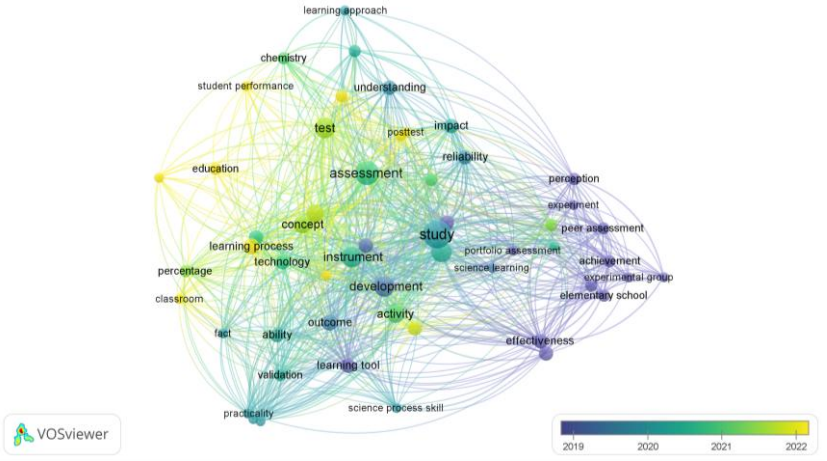

Figure 3. Overlay Visualization result

Figure 3 illustrates the evolution of research related to the development of assessment instruments in science education. Studies focusing on the development of assessment tools have

gained significant momentum from 2021 onwards, indicating a growing interest in designing more robust and relevant instruments for current educational needs. In contrast, earlier research—conducted primarily before 2019—focused heavily on conceptual understanding, learning outcomes, and experimental activities. This earlier emphasis is likely attributed to the pre-COVID-19 context, where hands-on experimental learning was more feasible, and learning goals were predominantly centered on content mastery.

However, recent shifts in educational policy—particularly the transition from the 2013 Curriculum to the Merdeka (Independent) Curriculum in Indonesia—have prompted a reorientation of learning objectives. The new curriculum emphasizes not only conceptual knowledge but also the development of students' character and competencies through the Pancasila Student Profile, which includes values such as critical thinking and global diversity. As a result, there is now an urgent need for research to focus on the development of assessment instruments that measure these 21st-century competencies, particularly those related to critical thinking and intercultural awareness.

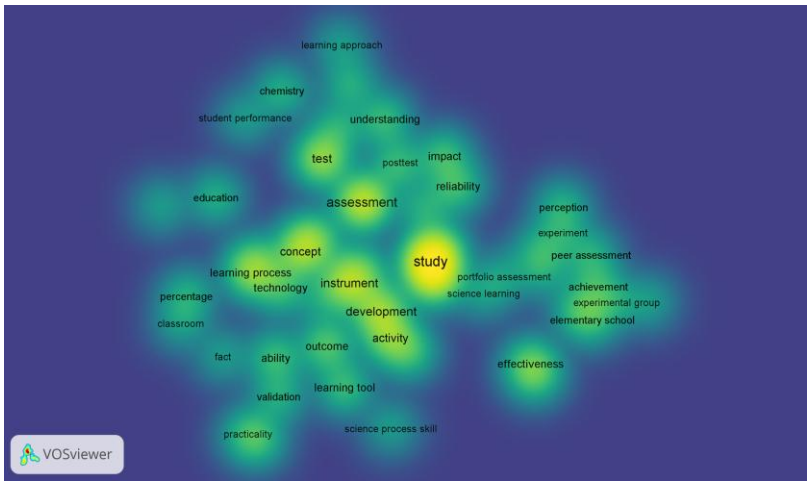

Figure 4. Density Visualization result

The density visualization, as shown in Figure 4, provides insight into the intensity and frequency of keyword usage across previous studies. The term "study" appears prominently in bright yellow, indicating its high occurrence and centrality in the reviewed literature. In contrast, the term "assessment" is displayed in green, signifying a moderate frequency and suggesting that research specifically focused on the development of assessment instruments remains limited.

Based on the visual patterns observed in Figures 2 through 4, it can be concluded that there is a significant gap in the literature concerning the development of assessment tools for 21st-century competencies, particularly critical thinking and global diversity. These findings highlight the urgent need for future research to focus on creating valid and reliable assessment instruments that address these essential skill areas, in alignment with modern educational goals.

3.2. Research instruments in science education from 2014 to 2024

Furthermore, the 20 selected studies on assessment in science education from 2014 to 2024, along with their citation data retrieved through Publish or Perish (PoP), are presented in Table 4. The data indicate that the most commonly utilized assessment format across these studies is the multiple-choice question (MCQ). This trend suggests a continued reliance on standardized, objective testing methods in science education assessment.

Table 4. Assessment research in science education from 2014 to 2024

No.	Author, Year	Education level	Number of Research subjects	Measured variables	Field of science	Kinds/ types of assessment	Research findings
1	(Hwang, Hung, & Chen, 2014)	Elementary school	167 sixth-grade students	Learning achievements, motivations and problem-solving skill	Integrated science	Open-ended	Most of the students assumed that open-ended assessment is an effective learning strategy to help them improve their learning status in "in-depth thinking", "motivations" and "creativity"
2.	(Chang, Hsu, & Jong, 2020)	Elementary school	Fifth-grade students	Geological knowledge, environment awareness, and earth science	Integrated science	Virtual reality integrated assessment	Students use a higher-level peer assessment learning approach while participating in VR design activities. Peer assessment boosts students' efficacy and propensity for critical thinking in addition to raising academic accomplishment
3.	(Parmiti, Rediani, Antara, & Jayadiningrat, 2021)	Elementary school	151 students	Scientific attitudes and science process	Integrated science	Not clearly stated	Science process competencies and scientific attitudes can be assessed through the evaluation of projects that depend local culture. The test and scientific attitude questionnaire exhibits an average Aiken validity coefficient of 0,85, according to the results. The test's reliability coefficient was 0,882, compared to the scientific attitude questionnaire's reliability coefficient was 0,0907, as calculated by the Cronbatch's Alpha formula.
4.	(Park, Liu, & Waight, 2017)	Senior High School	1230 students	Understanding of Chemistry	Chemistry of matter and energy	Connected Chemistry as Formative Assessment (CCFA)	CCFA supported the students learning process by making a connection between three chemistry domains: microscopic, sub microscopic, and representation of subject in chemistry of matter and energy.
5.	(Gunay & Ogan-Bekiroglu, 2014)	Senior High School	58 students	Research of the outcomes of physics learning and attitudes	Physics	Portfolio assessment	The evaluation of a student's portfolio serves as a teaching tool in along with providing insight into their cognitive capacities. Essentially, portfolio evaluation improves physics comprehension. This evaluation provides empirical evidence for the beneficial effects of portfolio assessment on learning and supports the assertion it extends students' knowledge. The study's findings demonstrate the students' initial experiences with portfolio evaluation do not lessen their favourable opinions of physics. Students' attitudes toward physics will therefore improve when portfolio assessment is incorporated into the curriculum.

6.	(Nsabayezu, Iyamuremye, Mbonyiryivuze, Niyonzima, & Mukiza, 2023)	Senior High School	50 participants	Formative assessment	Organic chemistry	Digital multiple choice	The use of digitally based formative evaluations, which enrich their personalized learning experiences and increase information retention, is well-received by students. The pupils showed how excited they were to study organic chemistry.
7.	(Isatunada & Haryani, 2021)	Junior High School	27 Junior High Schools in Semarang	Problem-solving ability and student's creativity	Integrated science of – global warming	-	Develop the learning assessment using STEM to measure students' problem-solving skills and creativity.
8.	(Amini & Fitria, 2019)	Elementary school	-	Learning outcomes	-	Performance assessment	Performance assessment instruments for elementary school students are valid and practical used in the leaning process.
9.	(Novitasari, Wulan, & Utari, 2018)	Senior High School	35 students	Scientific information literacy skills	-	Portfolio assessment in essay questions	Portfolio assessment in essay questions is valid and reliable based on the annates analysis and instruments indicating students could not evaluate information critically and competently at rate of 87%
10.	(Ligabo, Silva, Carvalho, Rodrigues, & Rodrigues, 2023)	Senior High School	25 students	Knowledge of organism diversity.	Biology matter of organism diversiy	Map concept assessment	In a quest to provide relevant learning experiences, concept map-based assessments measuring knowledge of organisms can be implemented.
11.	(Mustikasari, Yulianti, Hamimi, Affriyenni, & Lutfiani, 2021)	Junior High School	-	Concept understanding and self-efficiency	Integrated science on vibration waves and sound	-	Self-efficacy and conceptual knowledge assessment development validation findings average 97.8% and adhere to the 'excellent' criterion (97.9%).
12.	(Irma, Kusairi, & Yuliati, 2023)	Senior High School	30 students	Science creativity	Physics matters of static fluids	Essay questions	The development of an instrument to quantify scientific creativity provides takes into account indications of originality, fluency, elaboration, and adaptability, has a reliability of 0.610 and is deemed valid based on construct and content validity.
13.	(Putra, Fauzi, & Ratnawulan, 2019)	Junior High School	-	Learning outcomes	Integrated science on heat and its transfer	-	Valid instruments assessments at 86,8%, the average of instruments' effectiveness at 86,4%
14.	(Tarmo, 2022)	Senior High School	85 students	Learning achievements	Biology	Open-ended	Higher academic attainment in the experimental group is probably a result of assessment being incorporated into the learning process. This study advances our knowledge of how educators might use assessment strategies to incorporate learning into

							regular lessons in classrooms with limited resources, maximizing the potential of assessment to improve learning and elevate expectations.
15.	(Ayu, Widyaningsih, Ketut Dibia, & Sudana, 2020)	Elementary School		Learning outcomes	Integrated science on the beauty of my country's diversity	Multiple choice	With a validity score of 1.00, the assessment instrument created for students in elementary schools to assess their learning outcomes in scientific education was determined to be valid; its reliability, at 0.66, is regarded as high.
16.	(Yakob et al., 2023)	Senior High School	118 students	Scientific knowledge	Chemistry on chemical equilibrium	Virtual laboratory assessment	Assessment instruments based on a virtual laboratory can be used by the teacher to measure students' scientific knowledge
17.	(Dermawan, Wardani, Kurniawati, & Pranoto, 2021)	Elementary school	204 students	Science knowledge is based on HOTS (higher order thinking skills)	Environment with the theme of our friends	Multiple choice	Practical, valid and reliable instrument assessment based on HOTS to measure the science knowledge can be used in science education
18.	(Makahinda & Mawuntu, 2023)	Senior High School	-	Concept understanding	Thermodynamic physics	Portfolio assessment	Portfolio assessment can be used to measure the task, and research has shown that using portfolio in the topics of dynamics produces students' high learning outcomes.
19.	(Mutmainah & Muchlis, 2022)	Senior High School	39 students	Cognitive assessment outcomes	Chemistry topics, chemical bonds	Multiple choice	Instrument development can be used as an assessment in the concept of chemical bonds. The results of this study provide empirical evidence that student learning outcomes can be improved through Assessment for Learning.
20.	(Adimayuda et al., 2021)	Senior High School	34 students	Concept understanding	The physics concepts of momentum and impulse	A written test students' worksheet	Worksheet test shown the average percentage of students ability in understanding of the concept of two-dimensions collision at 67,14%. The data showed that students' learning achievements is in a good category. The results of this study provide empirical evidence that student learning outcomes can be improved through Assessment for Learning.

The analysis presented in Table 4 reveals a significant trend in the use of multiple-choice questions (MCQs) as the dominant form of assessment in science education from 2014 to 2024. While other formats, such as essays and open-ended questions, are also utilized, they appear less frequently in the reviewed literature. Each of these assessment types has unique strengths and limitations that impact their effectiveness in measuring various learning outcomes.

MCQs are widely used due to their objectivity, efficiency, and scalability. They allow for quick and standardized scoring, which can be conducted automatically using digital tools—making them

particularly suitable for large-scale testing environments. MCQs are also time-efficient, enabling the assessment of a broad range of content within a limited time frame (Murti, Wiyanto, & Hartono, 2018).

However, one of the key limitations of MCQs is their potential to encourage surface-level learning if not designed thoughtfully. The quality of MCQs greatly depends on the skill of the item writer. While it is possible to create higher-order questions that assess application, analysis, or evaluation, poorly constructed items tend to focus only on factual recall. Furthermore, MCQs allow for guessing, which may inflate scores and fail to accurately represent students' true understanding (Mustikasari, Zulkardi, & Aisyah, 2010).

Essay-type questions require students to construct responses using their own language based on conceptual understanding. These questions are advantageous in assessing higher-order thinking, such as reasoning, justification, synthesis, and reflection. Essays allow students to demonstrate depth of understanding and the ability to articulate their thoughts clearly.

Nevertheless, the use of essay questions presents several challenges. Scoring is often subjective, as it depends on the evaluator's interpretation, which may lead to inconsistency and bias. Additionally, manual grading is time-consuming, particularly in large classes. Another drawback is that essay questions may only address a limited portion of the curriculum, making it difficult to assess students' overall mastery of the subject matter (Ruslan & Santoso, 2013).

Open-ended questions are designed to accept multiple correct responses and encourage students to engage with problems in creative and diverse ways. This type of assessment is particularly effective for evaluating students' critical thinking, problem-solving skills, and their ability to interpret and analyze complex scenarios. Open-ended formats also provide insights into students' cognitive processes, allowing teachers to detect misconceptions or gaps in reasoning (Mustikasari et al., 2010).

The primary strength of open-ended questions lies in their capacity to foster student creativity and flexible thinking. They offer opportunities for learners to propose various problem-solving approaches, thus supporting the development of 21st-century competencies. However, similar to essays, these questions can be time-consuming to grade and may require clearly defined rubrics to ensure consistent and fair assessment.

The findings from this study highlight a persistent reliance on cognitive assessments, particularly those targeting content recall and conceptual understanding. While such assessments have value—especially in standardizing student performance—they do not sufficiently address broader educational goals related to 21st-century skills, such as critical thinking, creativity, collaboration, and global competence (Redhana, 2019).

As education systems evolve toward more holistic learning frameworks—such as the Merdeka Curriculum in Indonesia, which promotes the Pancasila Student Profile—there is a pressing need to diversify assessment methods (Sari, 2024). Current assessment practices must shift from focusing solely on knowledge acquisition to evaluating students' readiness to navigate complex, real-world challenges. This includes measuring competencies like critical thinking and intercultural awareness, which are essential for success in the 21st century (Anwar & Uman, 2020).

To address these emerging needs, future research should prioritize the design and validation of innovative assessment instruments. These tools should integrate multiple assessment formats and be aligned with both content standards and key competencies. In particular, assessment development efforts must consider alternative formats that foster deep thinking, real-world application, and reflective learning processes.

Additionally, leveraging technology-enhanced assessments may help overcome the limitations of traditional methods. For example, digital platforms can support adaptive testing, automated feedback, and data analytics to offer more personalized and meaningful assessment experiences.

In summary, while MCQs remain the most frequently used assessment format in science education, their limitations underscore the need for more balanced approaches. Essay and open-ended questions offer valuable alternatives for assessing higher-order thinking but require greater investment in design and scoring processes. Moving forward, the development of comprehensive, skills-based assessments will be critical to aligning science education with the demands of 21st-century learning.

4. CONCLUSION

Based on the findings of this study, it is evident that assessment practices in science education from 2014 to 2024 have predominantly focused on measuring conceptual understanding, learning outcomes, and experimental knowledge, especially in studies conducted prior to 2019. This is reflected in the widespread use of multiple-choice assessments, which primarily target cognitive skills and content recall. However, such formats are limited in their ability to evaluate higher-order thinking, and they pose issues such as answer guessing, which can compromise the validity of the results. Notably, there is a lack of research and development of assessment instruments aimed at measuring critical thinking skills and global diversity, two essential components of 21st-century learning. This represents a significant gap in the literature and highlights the limitations of current assessment trends, which remain heavily reliant on traditional cognitive-focused tools. As a limitation, this study is based solely on a systematic literature review and relies on bibliometric analysis, which may exclude relevant studies not indexed or captured by the selected databases. Therefore, future research should focus on the design and validation of innovative, skill-integrated assessment instruments that address broader educational goals. In particular, further empirical studies are needed to develop and test alternative assessment formats—such as performance tasks, open-ended responses, and technology-enhanced assessments—that more effectively measure critical thinking, creativity, and intercultural competencies in science education.

Acknowledgments: The authors would like to thank the Indonesian Ministry of Education, Culture, Research, and Technology for funding this study through a *Program Magister Menuju Doktoral untuk Sarjana Unggul* (PMDSU) scheme with grant number 1076.1/UN27.22/PT.01.03/2024.

REFERENCES

Adimayuda, R., Sari, L., Ismail, A., Amalia, I. F., Gumilar, S., & Samsudin, A. (2021). Assessing two-dimensional collisions second-years students using PhET simulations on momentum and impulse concept. *Journal of Physics: Conference Series*, *1869*(1). https://doi.org/10.1088/1742-6596/1869/1/012204

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, *71*(2), 103–112. https://doi.org/10.4097/kjae.2018.71.2.103

Amelia, D., Mutmainah, I., Indriyani, I., & Carmenette, N. S. (2023). *The Study of Assessment in the Merdeka Curriculum at Elementary School*. *7*(2), 121–128.

Amini, R., & Fitria, Y. (2019). The development of performance assessment based on integrated model on static electrical in elementry school. *Journal of Physics: Conference Series*, *1317*(1). https://doi.org/10.1088/1742-6596/1317/1/012167

Anwar, & Uman. (2020). Transformative Education: Emphasizing 21st Century Skills and Competencies in The Independent Learning Curriculum. *AIM: Journal of Islamic Education Management*, *1*(1), 1–16.

Ayu, I., Widyaningsih, T., Ketut Dibia, I., & Sudana, D. N. (2020). Instrument Appropriateness On The Assessment Of Fourth Grade Science Learning Outcomes In Elementary Schools. *International Journal of Elementary Education*, *4*(4), 472–480. Retrieved from https://ejournal.undiksha.ac.id/index.php/IJEE

Bernard, R. M., Zhang, D., Abrami, P. C., Sicoly, F., Borokhovski, E., & Surkes, M. A. (2008). Exploring the structure of the Watson-Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, *3*(1), 15–22. https://doi.org/10.1016/j.tsc.2007.11.001

Byusa, E., Kampire, E., & Mwesigye, A. R. (2022). Game-based learning approach on students' motivation and understanding of chemistry concepts: A systematic review of literature. *Heliyon*, *8*(5), e09541. https://doi.org/10.1016/j.heliyon.2022.e09541

Chang, S. C., Hsu, T. C., & Jong, M. S. Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers and Education*, *146*, 103758. https://doi.org/10.1016/j.compedu.2019.103758

Dermawan, D. D., Wardani, S., Kurniawati, Y., & Pranoto, S. (2021). Development Using Higher Order Thinking Skills Assisted by Quizizz Applications in Science Learning. *Proceeding ISET (2021) Universitas Negeri Semarang*, *7*, 89–97.

Desriani, Takunas, R., & Nadira, S. (2021). Implementasi Essay Test dalam Meningkatkan Wawasan Peserta Didik pada Pembelajaran Pendidikan Agama Islam di SMA Negeri 6 Sigi. *Jurnal Al-Tawjih*, *2*(2), 155–182.

Dettweiler, U. (2019). The rationality of science and the inevitability of defining prior beliefs in empirical research. *Frontiers in Psychology*, *10*(AUG), 1–4. https://doi.org/10.3389/fpsyg.2019.01866

Gunay, A., & Ogan-Bekiroglu, F. (2014). Impact of portfolio assessment on physics students' outcomes: Examination of learning and attitude. *Eurasia Journal of Mathematics, Science and Technology Education*, *10*(6), 667–680. https://doi.org/10.12973/eurasia.2014.1227a

Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research and Development*, *62*(2), 129–145. https://doi.org/10.1007/s11423-013-9320-7

Irma, Z. U., Kusairi, S., & Yuliati, L. (2023). Strem Pbl With E-Authentic Assessment: Its Impact To Students' Scientific Creativity on Static Fluid. *Jurnal Pendidikan IPA Indonesia*, *12*(1), 80–95. https://doi.org/10.15294/jpii.v12i1.40214

Isatunada, A., & Haryani, S. (2021). Development of Science Learning Tools using the STEM Approach to Train Problem Solving Ability and Students Activeness in Global Warming Material. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 363–375. https://doi.org/10.24815/jpsi.v9i3.19599

Junedi, B., Mahuda, I., & Kusuma, J. W. (2020). Optimalisasi keterampilan pembelajaran abad 21 dalam proses pembelajaran pada Guru MTs Massaratul Mut'allimin Banten. *Transformasi: Jurnal Pengabdian Masyarakat*, *16*(1), 63–72. https://doi.org/10.20414/transformasi.v16i1.1963

Kaptan, K., & Timurlenk, O. (2012). Challenges for Science Education. *Procedia - Social and Behavioral Sciences*, *51*, 763–771. https://doi.org/10.1016/j.sbspro.2012.08.237

Kwok, S. (2018). Science Education in the 21st Century. *Nature Astronomy*, *2*(7), 530–533. https://doi.org/10.1038/s41550-018-0510-4

Lame, G. T. (2019). Systematic Literature Reviews : An Introduction. *International Conference on Engineering Design*, (July). https://doi.org/10.1017/dsi.2019.169

Ligabo, M., Silva, F. C., Carvalho, A. C. da S. A., Rodrigues, D., & Rodrigues, R. C. L. B. (2023). Practical way to apply fourth-generation assessment tools integrated into creating meaningful learning experiences in biology at high school. *Evaluation and Program Planning*, *96*(August 2022), 102155. https://doi.org/10.1016/j.evalprogplan.2022.102155

Makahinda, T., & Mawuntu, V. J. (2023). Development of Thermodynamics Learning With Empirical Approach and Portfolio Assessment Techniques. *Studies in Learning and Teaching*, *4*(2), 285–295.

https://doi.org/10.46627/silet.v4i2.263

Murti, Wiyanto, & Hartono. (2018). Studi Komparasi antara Tes Testlet dan Uraian dalam Mengukur Hasil Belajar Kognitif Siswa Kelas XI SMA Negeri 1 Gombong. *Unnes Physics Education Journal*, *7*(1), 23–31.

Mustikasari, V. R., Yulianti, E., Hamimi, E., Affriyenni, Y., & Lutfiani, H. (2021). Development of lesson plan learning cycle 7E models integrated web formative assessment and self efficacy. *AIP Conference Proceedings*, *2330*(March). https://doi.org/10.1063/5.0043378

Mustikasari, Zulkardi, & Aisyah, N. (2010). Pokok Bahasan Bilangan Pecahan. *Jurnal Pendidikan Matematika*, *4*(1).

Mutmainah, S., & Muchlis, M. (2022). Implementation of assessment for learning to improve students' cognitive learning outcomes in the concept of chemical bonding. *Jurnal Pijar Mipa*, *17*(2), 217–223. https://doi.org/10.29303/jpm.v17i2.3308

Nieto, S. (2017). Re-imagining multicultural education: new visions, new possibilities*. *Multicultural Education Review*, *9*(1), 1–10. https://doi.org/10.1080/2005615X.2016.1276671

Novitasari, D., Wulan, A. R., & Utari, S. (2018). Profile of information literacy on the 21 st century through implementation of portfolio assessment. *Universitas Pendidikan Indonesia*, *3*, 509–513.

Nsabayezu, E., Iyamuremye, A., Mbonyiryivuze, A., Niyonzima, F. N., & Mukiza, J. (2023). Digital-based formative assessment to support students' learning of organic chemistry in selected secondary schools of Nyarugenge District in Rwanda. In *Education and Information Technologies* (Vol. 28). Springer US. https://doi.org/10.1007/s10639-023-11599-7

Park, M., Liu, X., & Waight, N. (2017). Development of the Connected Chemistry as Formative Assessment Pedagogy for High School Chemistry Teaching. *Journal of Chemical Education*, *94*(3), 273–281. https://doi.org/10.1021/acs.jchemed.6b00299

Parmiti, D. P., Rediani, N. N., Antara, I. G. W. S., & Jayadiningrat, M. G. (2021). The effectiveness of local culture-integrated science learning through project-based assessment on scientific attitudes and science process skills of elementary school students. *Jurnal Pendidikan IPA Indonesia*, *10*(3), 439–446. https://doi.org/10.15294/JPII.V10I3.31301

Putra, M., Fauzi, A., & Ratnawulan. (2019). Development of the science learning tools for junior high school based on group investigation with skill process approach in heat and the movement learning materials. *Journal of Physics: Conference Series*, *1185*(1). https://doi.org/10.1088/1742-6596/1185/1/012135

Redhana. (2019). Mengembangkan Keterampilan Abad Ke-21 Dalam Pembelajaran Kimia. *Jurnal Inovasi Pendidikan Kimia*, *13*(1), 2239–2253.

Roorda, N. (2022). Characteristics of science. *Omniconomics*, (January), 190–206. https://doi.org/10.4324/9781003306269-16

Ruslan, & Santoso. (2013). Pengaruh Pemberian Soal Open-Ended Terhadap Kemampuan Penalaran Matematis Siswa. *Jurnal Kreano*, *4*(2), 138–149.

Santos-meneses, L. F., & Drugova, E. A. (2023). Trends in critical thinking instruction in 21 st -century research and practice : Upgrading instruction in digital environments. *Thinking Skills and Creativity*, *49*(August), 101383. https://doi.org/10.1016/j.tsc.2023.101383

Sari, T. V. (2024). *The Implementation of The Merdeka Curriculum in Learning English in Senior High School : Case Study*. *10*(1), 28–35.

Solikah, A. A., Saputro, S., Yamtinah, S., & Masykuri, M. (2024). Research Trends in Group Investigation Learning Model for Critical Thinking Skills in Science Learning. *Jurnal Inovasi Pendidikan IPA*, *10*(1), 62–75. https://doi.org/10.21831/jipi.v10i1.70942

Songer, N. B., & Ruiz-Primo, M. A. (2012). Assessment and science education: Our essential new priority? *Journal of Research in Science Teaching*, *49*(6), 683–690. https://doi.org/10.1002/tea.21033

Stephen Jay Gould N. (1997). Nonoverlapping Magisteria. *Natural Science*, *106*(2). https://doi.org/10.53763/fag.2014.11.95

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*.

Tarmo, A. (2022). *Integrating assessment for learning into the teaching and learning of secondary school biology in Tanzania*. *12*(2), 239–265. https://doi.org/10.25656/01

Van Eck, N. J., & Waltman, L. (2021). Manual VOSviewer. *Univeristeit Leiden*, (January), 54.

Yakob, M., Sari, R. P., Hasibuan, M. P., Nahadi, N., Anwar, S., & Islami, R. A. Z. El. (2023). the Feasibility of Authentic Assessment Instrument Through Virtual Laboratory Learning and Its Effect on Increasing Students' Scientific Performance. *Journal of Baltic Science Education*, *22*(4), 631–640. https://doi.org/10.33225/jbse/23.22.631

Yamin, M. (2019). Information technologies of 21st century and their impact on the society. *International Journal of Information Technology (Singapore)*, *11*(4), 759–766. https://doi.org/10.1007/s41870-019-00355-1