



Authentic assessment in elementary mathematics: A systematic review of task design, effectiveness, and research gaps

Dameis Surya Anggara  ^{1*}, Candra Abdillah ¹

¹ Elementary School Teacher Education Study Program, Faculty of Teacher Training and Education, Universitas Pamulang, South Tangerang, Indonesia 15417

Article info	Abstract
Keywords: authentic assessment, bibliometric analysis, performance-based tasks, primary mathematics education, PRISMA systematic review, student learning outcomes	Authentic assessment has emerged as an approach that bridges assessment practices with meaningful learning in elementary mathematics education. However, studies integrating task design, effectiveness, and research gaps remain limited. This study aims to systematically synthesize authentic assessment in elementary mathematics, focusing on task design, effectiveness, and future research directions. This research employs a Systematic Literature Review (SLR) guided by PRISMA. Articles were retrieved from Scopus using keywords related to authentic assessment and mathematics education. A total of 48 Scopus-indexed articles were identified, of which 15 were analyzed in depth according to inclusion criteria. Data were analyzed thematically and supported by bibliometric mapping using VOSviewer. The findings indicate that authentic assessment is commonly implemented through performance-based tasks, project-based assessment, portfolios, and rubric-based evaluation. These approaches are effective in enhancing students' conceptual understanding, critical thinking, problem-solving skills, and engagement in learning. Moreover, authentic assessment provides comprehensive diagnostic information to support differentiated instruction. However, several gaps remain, including limited teacher competence in assessment design, low integration of technology, a lack of culturally responsive assessment, and the scarcity of longitudinal studies. The present study contributes by offering a comprehensive synthesis and proposing directions for developing more adaptive, inclusive, and technology-integrated authentic assessment models in elementary mathematics education.

* Corresponding Author.

E-mail address: dosen13330@unpam.ac.id (Dameis Surya Anggara)

DOI: <http://dx.doi.org/10.33578/jpfkip.v15i4.1053-1069>

Received 31 March 2026; Received in revised form 2 June 2026; Accepted 7 July 2026

Available online 30 August 2026

e-ISSN 2598-5949 | p-ISSN 2303-1514 © The Authors.

1. Introduction

Assessment is a fundamental component of the learning process, serving as a primary source of information for evaluating instructional effectiveness and students' achievement of learning competencies. In the context of elementary mathematics education, assessment should not only measure students' mastery of concepts and procedural skills but also evaluate their critical thinking, problem-solving abilities, and the application of mathematical concepts in real-life situations (Black & Wiliam, 2009; J. Deal & G. Wismer, 2000; Morris et al., 2009). However, assessment practices in schools remain predominantly dominated by traditional objective-test approaches that emphasize lower-order cognitive skills while overlooking complex thinking processes and authentic learning contexts (Hu et al., 2025; Mitchell & McMillan, 2018; Saher et al., 2022). This condition reflects a mismatch between current assessment practices and the demands of contemporary mathematics education, which emphasizes the development of twenty-first-century competencies.

In response to these limitations, authentic assessment has emerged as an alternative approach that emphasizes the connection between assessment and real-world activities. Authentic assessment is implemented through contextual tasks, such as performance assessments, projects, and portfolios, requiring students to apply their knowledge and skills in meaningful situations (Vlachopoulos & Makri, 2024; Wiewiora & Kowalkiewicz, 2019). This approach is grounded in constructivist learning theory, which views learning as an active process of knowledge construction through meaningful experiences (Darling-Hammond, 2014; Pellegrino et al., 2016; Chudowsky, & Glaser, 2001). Numerous empirical studies have demonstrated that authentic assessment enhances learning motivation, deepens conceptual understanding, and promotes higher-order thinking skills as well as the transfer of knowledge to real-world contexts (Darling-Hammond, 2014; Gulikers et al., 2004; OECD, 2019).

Despite its strong theoretical and empirical foundations, authentic assessment in elementary mathematics education continues to face significant challenges. Teachers often encounter difficulties in designing genuinely authentic tasks, developing valid and reliable assessment instruments, and integrating cognitive, metacognitive, social, and affective competencies simultaneously (Torrance & Pryor, 2001; Andrade & Cizek, 2010; Brookhart, 2013). Furthermore, research on authentic assessment remains fragmented, often focusing on isolated aspects such as task design or learning outcomes without providing a comprehensive understanding of how these dimensions interact within elementary mathematics education. Other underexplored issues include the integration of technology into assessment practices, cultural responsiveness, and the long-term impact of authentic assessment on student learning development (Pellegrino et al., 2016; Solano-Flores & Nelson-Barber, 2001).

As previous studies have examined authentic assessment across various educational contexts, systematic investigations specifically mapping its implementation in elementary mathematics education remain limited. Therefore, the present study offers several novel contributions. First, it focuses specifically on authentic assessment within elementary mathematics education. Second, it integrates bibliometric analysis and thematic synthesis to map research trends, themes, and developments in the field. Third, it develops a typology of authentic assessment based on the characteristics of tasks used in elementary mathematics learning. Fourth, it constructs a research gaps map to identify underexplored areas within the literature. Lastly, it proposes a future research agenda to guide subsequent investigations on authentic assessment in elementary mathematics education. Accordingly, this study goes beyond merely synthesizing previous findings by providing a conceptual framework and a

roadmap for future research that may serve as a reference for researchers, practitioners, curriculum developers, and policymakers.

Given these circumstances, a systematic review is needed to comprehensively synthesize existing evidence. A Systematic Literature Review (SLR) is a methodological approach that enables the systematic, transparent, and replicable identification, evaluation, and synthesis of empirical evidence (Kitchenham, B., & Charters, 2007; Page et al., 2021). Guided by the PRISMA framework, the present study aims to map the landscape of authentic assessment in elementary mathematics education by focusing on three main aspects: the forms of authentic assessment tasks, their effectiveness, and existing research gaps, along with future research directions.

The urgency of this study is further reinforced by the increasing demand for twenty-first-century education, which emphasizes critical thinking, creativity, collaboration, and communication skills. Mathematics education is no longer solely concerned with procedural mastery but also with students' ability to use mathematics to understand and solve contextual problems in reflective and adaptive ways (Kilpatrick, Swafford, & Findell, 2001; OECD, 2019). Therefore, a comprehensive understanding of the design and implementation of authentic assessment is essential for improving the quality of mathematics teaching, learning, and assessment in elementary schools.

Based on the foregoing discussion, the present study focuses on identifying, evaluating, and synthesizing research on authentic assessment in elementary mathematics education. Accordingly, the study is guided by the following research questions:

1. What forms of authentic assessment tasks are implemented in elementary mathematics learning?
2. How effective is authentic assessment in elementary mathematics learning?
3. What research gaps and future research priorities can be identified regarding authentic assessment in elementary mathematics education?

2. Method

The present study employs a Systematic Literature Review (SLR) approach guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. It is to ensure that the identification, selection, and synthesis of the literature were conducted systematically, transparently, and replicably (Kitchenham, B., & Charters, 2007; Page et al., 2021). In addition to conducting a thematic synthesis of studies that met the inclusion criteria, this study integrated bibliometric analysis to map research trends and developments in authentic assessment within elementary mathematics education.

2.1 Literature search strategy

The literature search was conducted on March 5, 2026, using the Scopus database. This database was selected due to its broad coverage of scholarly publications and its relevance to the fields of education and assessment. The search strategy employs a combination of keywords and Boolean operators as follows: ("authentic assessment" OR "performance assessment" OR "project-based assessment" OR "portfolio assessment") AND ("mathematics education" OR "mathematical literacy") AND ("primary education" OR "elementary school" OR "primary school"). This search strategy was designed to capture variations in terminology used in the international literature on authentic assessment in the context of elementary mathematics education (Kerres & Bedenlier, 2020).

2.2 Inclusion and exclusion criteria

Inclusion and exclusion criteria were established to ensure the relevance and quality of the studies included in the review. The inclusion criteria comprised empirical studies on authentic assessment in elementary mathematics education, published in peer-reviewed journals, available in full text, and indexed in Scopus between 2016 and 2025. In contrast, conceptual papers, opinion articles, editorials, and studies conducted at educational levels other than elementary school were excluded from the analysis. The application of these criteria was essential to maintain the validity and consistency of the findings in the systematic review (Xiao & Watson, 2019). The detailed inclusion and exclusion criteria are presented in Table 1.

Table 1. Inclusion and exclusion criteria

Aspects	Inclusion Criteria	Exclusion Criteria
Publication Type	Empirical research articles (peer-reviewed journal articles)	Conceptual papers, opinion articles, editorials, book reviews
Publication Year	Articles published between 2016 and 2025	Articles published before 2016
Research Focus	Studies addressing authentic assessment, performance assessment, project-based assessment, or portfolio assessment in mathematics education	Studies not related to authentic assessment or mathematics education
Educational Level	Elementary school students (approximately 6–12 years old)	Secondary school, high school, or higher education studies
Source Quality	Scopus-indexed articles available in full-text format	Articles not indexed in reputable databases or unavailable in full text
Language	English-language articles	Articles published in languages other than English

2.3 PRISMA-based study selection process

The article selection process followed the PRISMA 2020 framework, consisting of four stages: identification, screening, eligibility, and inclusion. During the identification stage, 135 articles were retrieved from the Scopus database. Subsequently, 25 articles were excluded for being outside the study's scope, leaving 110 for screening. During screening, article titles, abstracts, and keywords were reviewed to determine their relevance to the research topic. A total of 27 articles were excluded because they did not meet the predefined keyword criteria. As a result, 83 articles proceeded to the full-text retrieval stage.

In the subsequent stage, 35 articles are excluded because they are not available in English. Following the retrieval and examination of full-text documents, 48 articles remain eligible for further assessment. At the eligibility stage, a comprehensive review of the full-text articles was conducted based on the established inclusion and exclusion criteria. A total of 33 articles are excluded because they are either inaccessible in full text or do not satisfy the analytical eligibility requirements. Consequently, 15 articles meet all criteria and are included in the final synthesis of this systematic literature review. This selection process ensures that the studies included in the review are highly relevant, of high quality, and aligned with the focus of authentic assessment in elementary mathematics education. The study selection procedure is illustrated in Figure 1.

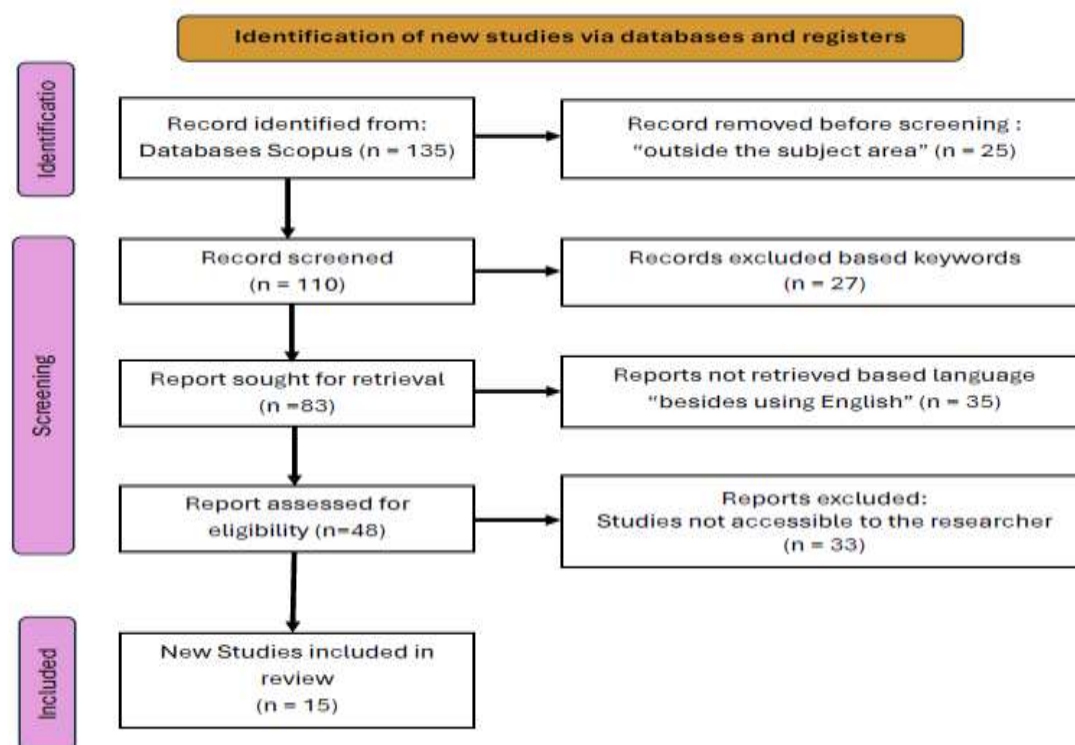


Figure 1. PRISMA 2020 flowchart

2.4 Data extraction and analysis

Data from the selected studies were systematically extracted to obtain essential information, including research characteristics, assessment instruments, task design, and key findings. This extraction process facilitated comparative analysis and cross-study synthesis. Subsequently, data analysis was conducted using a bibliometric approach with VOS viewer software to identify keyword co-occurrence patterns, map research networks through network visualization, and analyze topic density and research trends through density visualization (Donthu et al., 2021; van Eck & Waltman, 2010). This approach enables a systematic visualization of the knowledge structure and provides a comprehensive overview of the development of research on authentic assessment. In addition, a thematic synthesis of the selected studies was conducted to identify recurring themes related to the design, effectiveness, and research gaps in authentic assessment tasks in elementary mathematics education. The combination of bibliometric analysis and thematic synthesis enable both a macro-level mapping of the research landscape and an in-depth understanding of the empirical findings reported in the literature.

3. Results

3.1 Overview

3.1.1 Distribution of publication years

A total of 48 articles were identified as relevant to the research topic; however, only 15 were categorized as highly relevant. The present study also analyzes the development of authentic assessment in elementary school mathematics over the period 2016–2025, as illustrated in Figure 2.

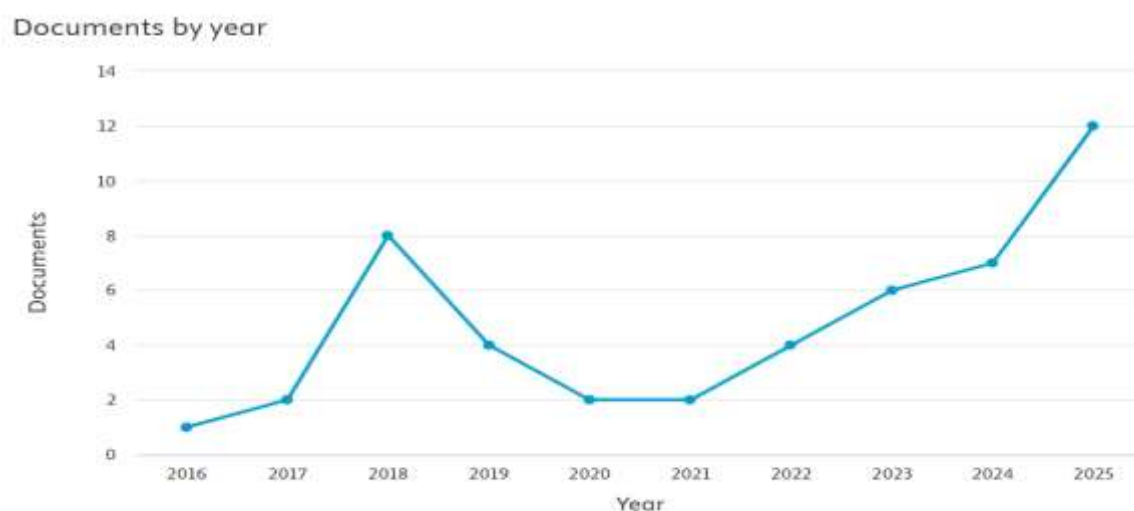


Figure 2. Distribution of publication years for authentic mathematics assessment in elementary schools

Based on the graph showing the number of documents per year for the period 2016–2025, there is a fluctuating pattern of publication productivity, with an overall increasing trend in recent years. In the early phase (2016–2017), the number of publications is relatively low, followed by a significant increase in 2018. This is subsequently followed by a decline in 2019, reaching its lowest point during 2020–2021. Entering 2022, the number of publications began to increase consistently, reaching its peak in 2025. Overall, this pattern indicates that, despite fluctuations in the earlier period, research in this field has shown a positive growth trend, particularly after 2021. This reflects the growing urgency and relevance of authentic assessment in supporting more meaningful mathematics learning at the elementary school level. The corresponding visualization is presented in Figure 3.

3.1.2 Country of origin of the studies

The distribution of publications across countries reveals varying levels of contribution to the development of this research area. Indonesia ranks first with 10 publications, indicating the strongest contribution to the field. The United Kingdom and the United States follow with six publications each, while South Africa contributes four publications and Australia three. Canada, Ireland, Italy, Luxembourg, and Malaysia contribute two publications. This pattern demonstrates that although research contributions originate from diverse regions, productivity remains concentrated in several countries, with Indonesia emerging as the most prominent contributor within the analyzed dataset. A comprehensive summary of the reviewed studies, including article summaries, key characteristics, research designs, participants, data collection techniques, and major findings, is presented in Appendix 1.

Documents by country or territory

Compare the document counts for up to 15 countries/territories.

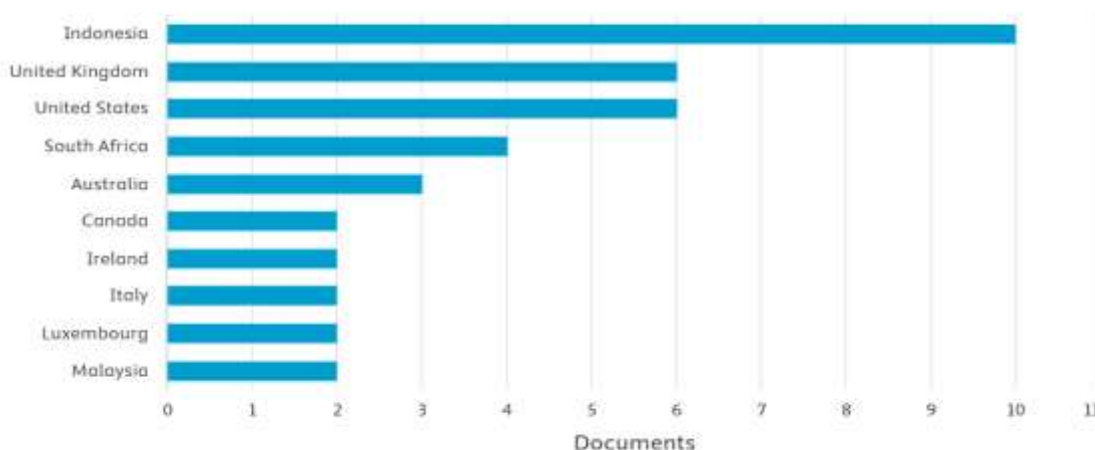


Figure 3. Country of origin of studies on authentic assessment in elementary school mathematics

3.2 Thematic findings

The results of the bibliometric analysis using VOSviewer software reveal a structured pattern of keyword relationships, forming several thematic clusters in research on authentic assessment in elementary school mathematics education. This network visualization provides an overview of the conceptual relationship patterns that have developed within this research field.

In the visualization, the size of each node represents the frequency of a term in the analyzed publications. Therefore, the larger the node, the higher the term's frequency. Meanwhile, the connecting lines (links) between nodes indicate the strength of relationships or the level of association among keywords, based on their co-occurrence frequency. Figure 4 presents the Network Visualization Results.

Referring to Figure 4, research on authentic assessment in elementary school mathematics education is grouped into four main clusters: (1) learning and authentic assessment, (2) mathematics and student characteristics, (3) implementation context in elementary schools, and (4) assessment instrument development. These four clusters illustrate an interconnected conceptual structure within the study of authentic assessment.

The red cluster represents a strong relationship between authentic assessment and learning, teacher feedback, educational measurement, problem-based learning, and experiential learning. This finding confirms that authentic assessment is closely integrated with active, student-centered learning processes. In this context, teachers play a key role in designing, implementing, and reflecting on assessment practices. Furthermore, the presence of feedback underscores the urgency of constructive feedback in supporting the continuous development of students' understanding and skills.

The green cluster reveals that mathematics is associated with children, math anxiety, decision making, and mathematics assessment. This indicates that mathematics learning is not limited to cognitive aspects but also involves students' affective and psychological dimensions. The presence of math anxiety underscores that it can influence students' decision-making processes and learning outcomes. Therefore, the assessment design should balance cognitive aspects with students' psychological conditions.

The blue cluster represents the implementation context of authentic assessment in elementary classrooms, indicated by the connections among students, primary school, teaching, classroom assessment, and assessment. This cluster emphasizes that authentic assessment must be meaningfully integrated into daily classroom practices. Meanwhile, the yellow cluster focuses on the development of assessment instruments, particularly performance assessment, as shown by its association with assessment instruments, elementary school, and mathematics education. This highlights the importance of developing assessment tools that align with the characteristics of elementary mathematics learning.

Overall, the interconnections among these clusters indicate that authentic assessment in elementary school mathematics education is a complex, multidimensional system encompassing pedagogical, psychological, contextual, and technical aspects. The integration of learning processes, student characteristics, implementation contexts, and instrument development underscores that authentic assessment design must be holistic and adaptive to students' needs and instructional demands.

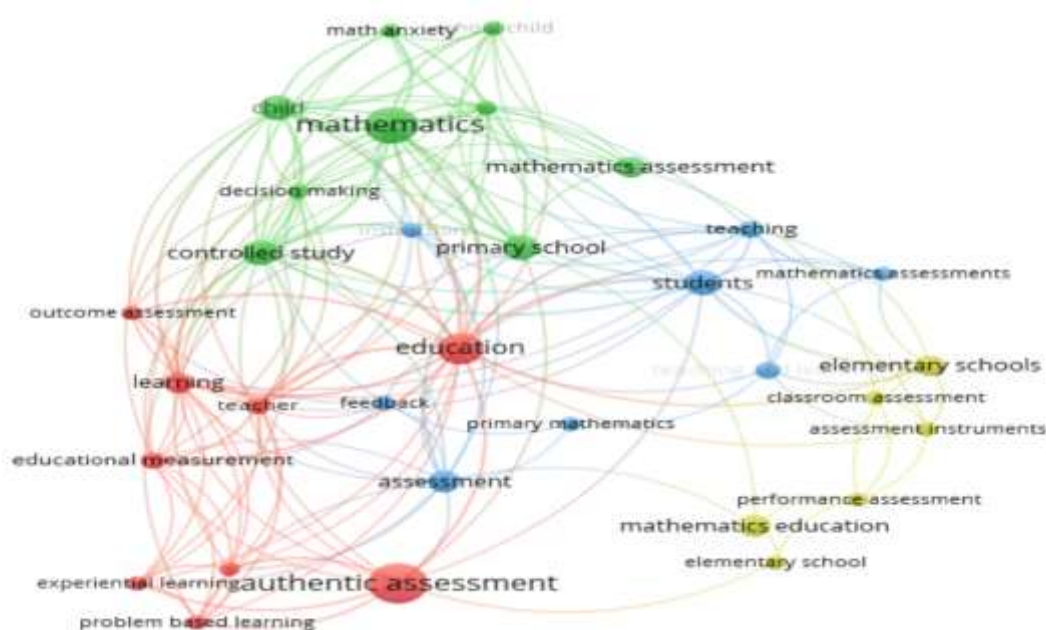


Figure 4. Network visualization results

Thus, this network map demonstrates that research on authentic assessment in elementary mathematics education is evolving toward a more comprehensive conceptual framework. A summary of the strengths, challenges, and limitations of the reviewed studies is presented in Table 2.

Table 2. Summary of strengths, challenges, and research limitations

Categories	Key Points
Strengths	Authentic assessment can holistically measure students' abilities, enhance conceptual understanding and numeracy, and provide diagnostic information for appropriate instructional interventions. It also promotes meaningful learning and the development of 21 st -century skills, and has the potential to reduce bias through innovations such as non-verbal assessment.

Categories	Key Points
Challenges	The implementation of authentic assessment still faces several challenges, including limited teacher understanding, a gap between policy and practice that remains dominated by traditional assessment, and the complexity of designing valid and reliable instruments. In addition, technology integration is not yet optimal, and cultural and language factors have not been fully incorporated into the assessment design.
Research Limitations	Research on authentic assessment is still dominated by short-term studies with limited scope, resulting in constrained generalizability and limited understanding of long-term impacts. Moreover, numerous studies focus on instrument development without large-scale implementation testing, with limited multidimensional integration. Cultural- and technology-based studies remain scarce, while the dominance of standardized assessment may introduce bias and fail to fully represent students' mathematical abilities.

4. Discussion

4.1 Authentic assessment tasks in elementary mathematics education

Authentic assessment in elementary mathematics education represents a paradigm shift from traditional outcome-oriented assessment toward performance-based, contextual, and learning-centered assessment. This approach emphasizes students' active engagement in tasks that reflect real-world situations, enabling a more comprehensive evaluation of mathematical competencies, including knowledge, skills, thinking strategies, and problem-solving abilities (Vlachopoulos & Makri, 2024b; Yudha et al., 2019). Through authentic assessment, students' abilities are evaluated not only by their final answers but also by the reasoning processes and solution strategies they employ.

The implementation of authentic assessment is reflected in various forms of assessment tasks that extend beyond conventional written examinations. One of the most frequently adopted forms is performance-based assessment, which requires students to apply mathematical concepts in real-life contexts. Such tasks integrate conceptual understanding, procedural skills, critical thinking, and mathematical communication within meaningful learning activities (Darling-Hammond, 2014; Mcfeetors & Marynowski, 2021). In addition to performance tasks, project-based assessment is an essential means of evaluating students' ability to connect mathematical concepts to real-world phenomena over time.

The successful implementation of authentic assessment largely depends on the use of systematic and well-structured assessment instruments. Rubrics and observation checklists are widely used to objectively and consistently evaluate multiple dimensions of student performance. These instruments generally include assessment objectives, achievement indicators, performance criteria, and clearly defined achievement levels, thereby enhancing the validity and reliability of the assessment process (Tashtoush et al., 2025; Yudha et al., 2019). Consequently, assessment functions not only as a measurement tool but also as a mechanism for providing feedback that supports learning improvement.

Recent developments indicate that authentic assessment increasingly incorporates progression-based approaches designed to continuously and systematically monitor students' numeracy development. This approach provides teachers with more accurate information regarding students' learning progress and supports data-informed instructional decision-making (Balt et al., 2020; Wu et al., 2024). Furthermore, multidimensional diagnostic assessments have emerged as valuable tools for identifying various aspects of mathematical competence, including conceptual understanding, procedural fluency, adaptive reasoning, and strategic competence. Such approaches are particularly crucial for supporting inclusive education, as

they accommodate diverse student characteristics and learning needs (Black & Wiliam, 2009; Brendefur et al., 2017).

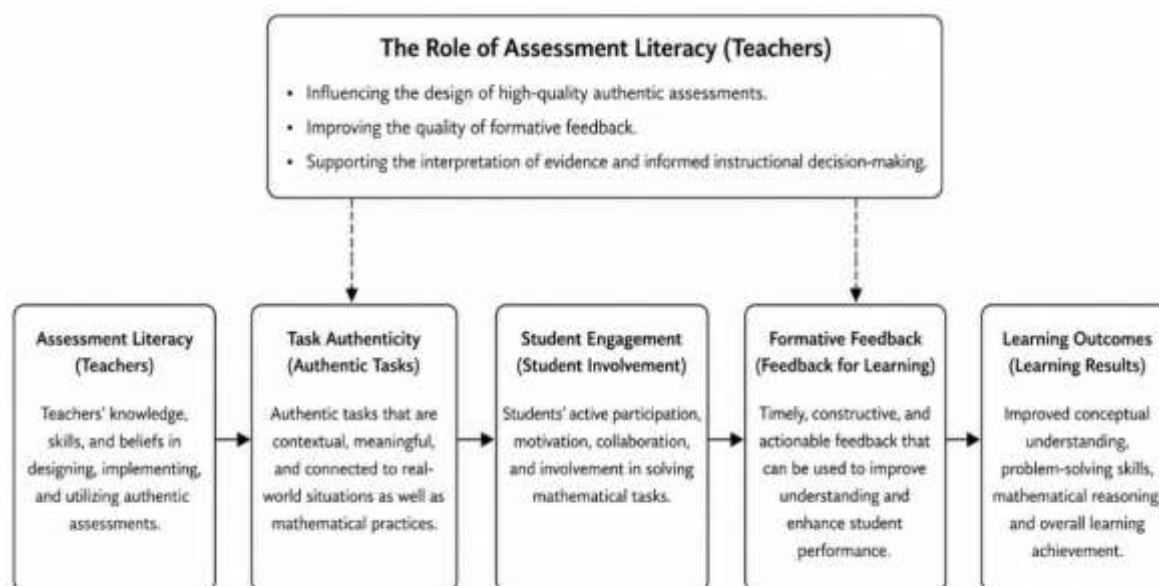


Figure 5. Conceptual framework of authentic assessment in elementary school mathematics

Global innovations in authentic assessment have also been characterized by the increasing use of digital technologies and non-verbal approaches. The integration of videos, interactive simulations, multimedia applications, and digital assessment platforms enables the development of more contextual tasks while reducing linguistic bias in assessment practices. These innovations have been shown to improve assessment accessibility for students from diverse social, cultural, and ability backgrounds (Greisen et al., 2018; Pellegrino et al., 2016). Therefore, authentic assessment in elementary mathematics has evolved into a dynamic, adaptive, and competency-oriented assessment system aligned with the demands of twenty-first-century education.

Based on the synthesis findings, authentic assessment tasks in elementary mathematics should no longer be viewed as isolated assessment components but rather as an integrated ecosystem that combines performance tasks, contextual projects, rubric-based observations, diagnostic assessments, learning progress monitoring, and digital technologies within a continuous assessment framework. The findings underscore that the effectiveness of authentic assessment is influenced not only by the types of tasks employed but also by the interaction among task authenticity, student engagement, assessment literacy, formative feedback, and learning outcomes. Authentic tasks promote meaningful student engagement in mathematical activities, while teachers' assessment literacy influences both task design and the quality of feedback. The interaction among these components ultimately contributes to improved mathematics learning outcomes. Consequently, this review extends previous studies that examined assessment forms separately by proposing a conceptual framework that explains the relationships among assessment design, learning processes, and learning outcomes in elementary mathematics education. The proposed conceptual framework is presented in Figure 5.

4.2 The effectiveness of authentic assessment in elementary school mathematics

A growing body of research indicates that authentic assessment is highly effective in enhancing the quality of mathematics learning in elementary schools. Its effectiveness is particularly evident in improving students' conceptual understanding through active engagement with tasks grounded in real-life contexts. Unlike traditional assessments that primarily emphasize procedural mastery and final outcomes, authentic assessment allows students to construct knowledge through exploration, reflection, and the application of mathematical concepts in meaningful situations. This process facilitates a deeper and more sustainable understanding of mathematical concepts (Balt et al., 2020; Barkatsas & McLaughlin, 2021).

The effectiveness of authentic assessment is also reflected in its ability to support meaningful learning by integrating assessment into everyday instructional practices. Based on the perspective of assessment as learning, assessment serves not merely as a means of measuring achievement but also as a tool through which students monitor, evaluate, and reflect on their learning progress. As a result, students become active learners engaged in self-regulated learning, while teachers gain richer information regarding students' competency development (Black & Wiliam, 2009; Mcfeetors & Marynowski, 2021).

In addition, authentic assessment has proven effective in providing comprehensive and continuous diagnostic information. Data generated through performance observations, projects, portfolios, and contextual tasks enable teachers to identify learning difficulties more accurately than conventional assessment methods. Such information can be used to design instructional interventions tailored to individual student needs, thereby supporting differentiated instruction in heterogeneous classrooms (Robiansyah et al., 2017; Vlachopoulos & Makri, 2024b). In this regard, authentic assessment serves not only as an evaluation tool but also as an evidence-based instrument for pedagogical decision-making. Authentic assessment also contributes to the development of twenty-first-century competencies that are increasingly essential in contemporary education. Through tasks that require problem-solving, investigation, group discussion, and communication of results, students develop critical thinking, creativity, collaboration, and communication skills in an integrated manner. These competencies not only support success in mathematics learning but also prepare students for future challenges in life and work (Darling-Hammond, 2014; Pellegrino et al., 2016).

Furthermore, authentic assessment promotes more equitable and inclusive educational practices. Studies indicate that contextual tasks, visual media, and digital technologies can reduce linguistic and cultural biases often associated with traditional assessments. Such approaches enable students from diverse social, cultural, and ability backgrounds to demonstrate their competencies through multiple forms of representation, resulting in more accurate and equitable assessment outcomes (Greisen et al., 2018; Pellegrino et al., 2016). Therefore, authentic assessment not only enhances mathematics learning but also strengthens the principles of equity and accessibility in elementary education.

Overall, the findings indicate that the effectiveness of authentic assessment should not be understood solely in terms of improved academic achievement, but rather in terms of its ability to integrate formative, diagnostic, and competency-development functions within a continuous assessment system. This perspective extends previous research that primarily focused on academic outcomes. The present review highlights that authentic assessment is effective because it fosters a learning ecosystem that supports conceptual understanding, data-informed pedagogical decision-making, twenty-first-century skill development, and inclusive, student-centered learning practices. Consequently, authentic assessment can be regarded as a strategic

component in transforming elementary mathematics education toward more adaptive, meaningful, and sustainable learning environments.

4.3 Research gaps and future research priorities

Despite being an authentic assessment, it has considerable potential to improve the quality of elementary mathematics education; the synthesis identified several research gaps warranting further investigation. First, a gap exists between the conceptual foundations and practical implementation of authentic assessment. While most studies emphasize the benefits of authentic assessment for conceptual understanding, problem-solving, and student engagement, classroom practices remain largely dominated by traditional written examinations (Syaifuddin, 2020; Yudha et al., 2019). This finding underscores that the primary challenge lies not in the lack of empirical evidence but in translating theory into classroom practice. Second, limited teacher assessment literacy consistently emerges as a major barrier across diverse educational contexts. Numerous teachers continue to perceive assessment primarily as an administrative requirement rather than as a tool to support learning (Chihodzi et al., 2023; Gaikwad et al., 2023). This condition indicates that the effectiveness of authentic assessment is strongly influenced by teachers' competencies in task design, rubric development, and assessment interpretation. Therefore, practice-oriented professional development programs aimed at strengthening teachers' assessment literacy should become a research priority (Black & Wiliam, 2009; Mirsanjari, 2025).

Third, the integration of technology into authentic assessment remains at an early stage of development. Notwithstanding several studies reporting positive outcomes associated with multimedia, digital assessments, and interactive platforms, empirical evidence regarding their effectiveness remains limited and sometimes inconsistent (Madland et al., 2024; Pellegrino et al., 2016). Future research should therefore investigate how technology can support authentic assessment from a pedagogical perspective rather than merely serving as a digital assessment tool. Fourth, research on culturally responsive assessment remains relatively scarce compared with studies focusing on performance assessment and diagnostic assessment. Yet, evidence reveals that cultural and linguistic contexts may significantly influence the validity of mathematics assessment (Ataide Pinheiro, 2025; Ramadhani et al., 2025). This gap highlights the need to develop assessment models that are more responsive to cultural diversity, language differences, and student characteristics. Fifth, most identified studies employed short-term research designs and therefore provide limited evidence regarding the sustained impact of authentic assessment on students' mathematical development. The scarcity of longitudinal studies restricts our understanding of the long-term effectiveness of authentic assessment (Darling-Hammond, 2014).

Based on these gaps, future research should focus on five main priorities: (1) developing adaptive and evidence-based digital authentic assessment models; (2) strengthening culturally responsive assessment to promote equity and inclusiveness in mathematics learning; (3) conducting longitudinal studies to evaluate the long-term effects of authentic assessment on student competency development; (4) designing practice-oriented programs to enhance teachers' assessment literacy; and (5) integrating performance assessment, diagnostic assessment, and portfolio assessment within a comprehensive assessment framework. These research priorities are expected to strengthen the theoretical foundations and broaden the implementation of authentic assessment in elementary mathematics education. A summary of the proposed future research agenda is presented in Table 3.

Table 3. Research agenda

No	Research Gaps		Recommended Methods	Participants		Variables		Expected Contributions
1	Digital assessment	authentic	Design-Based Research	Grades elementary students	4–6	Digital assessment, engagement		Digital assessment model
2	Culturally assessment	responsive	Mixed Methods	Multicultural students		Cultural achievement	relevance,	More equitable assessment
3	Longitudinal impact		Longitudinal Study	Grades elementary students	1–6	Authentic assessment, achievement growth		Evidence of long-term impact
4	Teacher literacy	assessment	Experimental Study	Elementary school teachers		Training, assessment quality		Enhanced teacher competence
5	Integrated model	assessment	Development Research	Teachers and students		Performance, diagnostic, assessment	portfolio	Comprehensive assessment model

5. Conclusion and implications

Based on this systematic literature review, authentic assessment in elementary mathematics education reflects a shift from traditional to performance-based, contextual, meaningful, and student-centered practices. Various forms of authentic assessment, including performance, portfolio, diagnostic, progression-based, and technology-enhanced assessments, provide more comprehensive evaluations of students' mathematical competencies. The synthesis reveals that authentic assessment contributes to the improvement of conceptual understanding, critical thinking, problem-solving skills, student engagement, and the provision of diagnostic information that supports differentiated and inclusive learning. This study contributes to the literature by providing an integrated synthesis of authentic assessment task types, effectiveness, research gaps, and future research directions in elementary mathematics education. Furthermore, the study offers a typology of authentic assessment, a research gaps map, and a future research agenda that may serve as valuable references for researchers, educators, and educational practitioners.

The implications of this study underscore that assessment should be positioned as an integral component of the learning process rather than merely a tool for measuring learning outcomes. Teachers need to strengthen their competencies in designing authentic tasks, developing valid and reliable assessment rubrics, and utilizing assessment results to provide continuous formative feedback. At the curriculum and policy levels, support is needed to implement diverse forms of authentic assessment and to integrate technology to enhance the quality, effectiveness, and accessibility of mathematics assessment.

6. Limitation

This study has several limitations. First, the review included only articles that met the established inclusion criteria and were published in English within the selected database, which may have resulted in the exclusion of relevant studies. Second, the number of studies included in the final synthesis was relatively limited, and the reviewed studies exhibited substantial methodological diversity. Consequently, the findings should be interpreted with caution. Future research should expand the scope of databases, include studies from diverse cultural contexts,

and conduct longitudinal investigations to examine the long-term effects of authentic assessment on the development of students' mathematical competencies.

Declaration and Statement

CRedit Authorship Contribution Statement

Dameis Surya Anggara: Conceptualization, Methodology, Data curation, Formal analysis, Writing original draft, Writing review & editing. Candra Abdillah: Methodology, Writing review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this manuscript.

Funding Statement

This research was funded by Universitas Pamulang through its internal research grant program.

Data Availability Statement

The data supporting the findings of this study are derived from publicly available published literature. The data extracted and analyzed in this systematic review are available from the corresponding author upon reasonable request.

Ethical Declaration

Ethical approval is not applicable to this manuscript because this study is a systematic review of previously published literature and does not involve human participants, animals, or the collection of primary data from individuals.

Informed Consent Statement

Informed consent was not applicable because this study is a systematic review of previously published literature and did not involve human participants or the collection of primary data from individuals.

Acknowledgement

The authors would like to express their sincere gratitude to Universitas Pamulang for providing institutional support for this research through its internal research grant program. The authors also appreciate the support provided by colleagues who contributed to the completion of this study.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with language editing and improving the clarity of the manuscript. All AI-assisted content was carefully reviewed, edited, and validated by the authors. The authors take full responsibility for the content, accuracy, and integrity of the final published manuscript.

References

- Andrade, H. L., & Cizek, G. J. (Eds.). (2010). Handbook of formative assessment. Routledge.
- Ataide Pinheiro, W. (2025). Mathematics as a Discursively Exclusionary Discipline to Queer Subjectivity: A Perspective Through Teaching Mathematics for Social Justice. Education Sciences, 15(9). <https://doi.org/10.3390/educsci15091116>
- Balt, M., Fritz, A., & Ehler, A. (2020). Insights Into First Grade Students' Development of Conceptual Numerical Understanding as Drawn From Progression-Based Assessments. Frontiers in Education, 5(80), 1–11. <https://doi.org/10.3389/educ.2020.00080>

- Barkatsas, T., & McLaughlin, P. (2021). Authentic Assessment and Evaluation Approaches and Practices in a Digital Era. *Authentic Assessment and Evaluation Approaches and Practices in a Digital Era*. <https://doi.org/10.1163/9789004501577>
- 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>
- Brendefur, J. L., Johnson, E. S., Thiede, K. W., Strother, S., & Severson, H. H. (2017). Developing a Multi-Dimensional Early Elementary Mathematics Screener and Diagnostic Tool : The Primary Mathematics Assessment. *Early Childhood Education Journal*, 0(0), 0. <https://doi.org/10.1007/s10643-017-0854-x>
- Chihodzi, B., Mwakapenda, W., & Ngulube, B. (2023). Ticks and crosses in primary mathematics assessments : What purpose do they serve ? *Pythagoras - Journal of the Association for Mathematics Education of South Africa*, 44(1), 1–13. <https://doi.org/10.4102/pythagoras.v44i1.647>
- Connor, B. R. O. (2024). Methodologies to reveal young Australian Indigenous students ' mathematical proficiency. *Mathematics Education Research Journal*, 36(2), 311–338. <https://doi.org/10.1007/s13394-023-00447-z>
- Darling-Hammond, L. (2014). One Piece of the Whole: Teacher Evaluation as Part of a Comprehensive System for Teaching and Learning. *American Educator*, 38(1), 4–13. <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1023870&site=ehost-live>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133(May), 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Gaikwad, S., Wadegaonkar, A., Mitra, G., & Chakravarty, D. (2023). Assessment Literacy, Current Assessment Practices and Future Training: Reflections of Teachers in Higher Education. *International Journal of Learning, Teaching and Educational Research*, 22(7), 1–29. <https://doi.org/10.26803/ijlter.22.7.1>
- Garcia, M. L. B., Santos, K. C. P., & Vistro-Yu, C. P. (2025). Comparing two psychometric approaches: The case of item analysis for a classroom test in mathematics. *International Journal of Education and Practice*, 13(1), 327–344. <https://doi.org/10.18488/61.v13i1.4060>
- Greisen, M., Hornung, C., Baudson, T. G., Muller, C., Martin, R., & Schiltz, C. (2018). Taking Language out of the Equation : The Assessment of Basic Math Competence Without Language. *Frontiers in Psychology*, 9(1076), 1–13. <https://doi.org/10.3389/fpsyg.2018.01076>
- Gulikers, J. T. M., Bostiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67–86. <https://doi.org/10.1007/BF02504676>
- Hortelano, J. C., & Prudente, M. (2025). Development of a questionnaire for understanding mathematics teachers ' situated learning practices in the Philippines. *Waikato Journal of Education*, 29(1), 24–39. <https://doi.org/10.15663/wje.a1064>
- Hu, A., Liu, Q., & Daniel, B. (2025). Digital Technologies in Authentic Assessment in Higher Education: A Systematic Literature Review and Narrative Synthesis. *SAGE Open*, 15(3). <https://doi.org/10.1177/21582440251357198>
- J. Deal, L., & G. Wismer, M. (2000). National Council of Teacher Mathematics, Principle and Standarts for School Mathematich, (USA:NCTM,2000). 55–65. <https://files.eric.ed.gov/fulltext/EJ893807.pdf>
- Kazima, M., Longwe, J., & Gobede, F. (2022). Improving teaching and learning mathematics in Malawi primary schools: a review of reforms, interventions, successes and challenges Mercy.

-
- Journal of Applied Learning & Teaching, 5(2), 81–89.
<https://doi.org/10.37074/jalt.2022.5.S2.3>
- Kerres, M., & Bedenlier, S. (2020). Systematic Reviews in Educational Research. In Systematic Reviews in Educational Research. <https://doi.org/10.1007/978-3-658-27602-7>
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report, Ver. 2.3 EBSE Technical Report. EBSE, January 2007, 1–57.
- Madland, C., Irvine, V., DeLuca, C., & Bulut, O. (2024). Technology-Integrated Assessment: A Literature Review. The Open/Technology in Education, Society, and Scholarship Association Journal, 4(1), 1–48. <https://doi.org/10.18357/otessaj.2024.4.1.57>
- Mcfeetors, P. J., & Marynowski, R. (2021). Generative Unit Assessment : Authenticity in Mathematics Classroom Assessment Practices. Education Sciences, 11(17), 1–22. <https://doi.org/10.3390/educsci11070366>
- Mirsanjari, Z. (2025). Enhancing assessment literacy in EAP instruction: the role of teacher development courses in overcoming systemic barriers. Language Testing in Asia, 15(1). <https://doi.org/10.1186/s40468-025-00368-7>
- Mitchell, K. M., & McMillan, D. E. (2018). A curriculum-wide assessment of writing self-efficacy in a baccalaureate nursing program. Nurse Education Today, 70(December 2017), 20–27. <https://doi.org/10.1016/j.nedt.2018.08.003>
- Morris, K. A., Hiebert, J., & Spitzer, M. S. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn? Journal for Research in Mathematics Education, 40(5), 491–529. <https://doi.org/10.5951/jresmetheduc.40.5.0491>
- 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. <https://doi.org/10.1136/bmj.n71>
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. Educational Psychologist, 51(1), 59–81. <https://doi.org/10.1080/00461520.2016.1145550>
- Pinheiro, W. A., Carrizales, D., Greenlees, L., Valle, F., Shive, E. C., & Phelps, R. (2025). Bilingual Teacher Candidates : Addressing Cultural Assumptions in Standardized Mathematics Assessment for Elementary Students Through Culturally Relevant Pedagogy. Education Sciences, 15(3), 313. <https://doi.org/10.3390/educsci15030313>
- Putra, Z. H., Rahmadhani, D., Noviana, E., & Hermita, N. (2022). Prospective elementary teachers ' attitude toward technology-based mathematics assessment Prospective elementary teachers ' attitude toward technology - based mathematics assessment. Journal of Physics: Conference Series, 2279(012007), 6–11. <https://doi.org/10.1088/1742-6596/2279/1/012007>
- Ramadhani, R., Soeharto, S., Arifiyanti, F., Prahmana, R. C. I., Saleh, A., & Lavicza, Z. (2025). Assessing quality and biases in ethnomathematics-based numeracy worksheets: A Many-Facet Rasch Model analysis. Social Sciences and Humanities Open, 12(December 2024), 101736. <https://doi.org/10.1016/j.ssaho.2025.101736>
- Robiansyah, S. T. U., Nanang, F., & Hidayat. (2017). Development Mathematic Assessment to Increase Mathematical Prerequisite Ability on The Student with Learning Disabilities in Inclusive Elementary Development Mathematic Assessment to Increase Mathematical Prerequisite Ability on The Student with Learning. IOP Conf. Series: Materials Science and Engineering, 288(012129), 1–8. <https://doi.org/10.1088/1757-899X/288/1/012129>

- Saher, A. S., Ali, A. M. J., Amani, D., & Najwan, F. (2022). Traditional Versus Authentic Assessments in Higher Education. *Pegem Egitim ve Ogretim Dergisi*, 12(1), 283–291. <https://doi.org/10.47750/pegegog.12.01.29>
- Sayac, N., & Veldhuis, M. (2022). Mathematics Assessment Practices of Primary School Teachers in France. *International Journal of Science and Mathematics Education*, 20, 1595–1610. <https://doi.org/10.1007/s10763-021-10229-3>
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573. <https://doi.org/10.1002/tea.1018>
- Syaifuddin, M. (2020). Implementation of authentic assessment on mathematics teaching: Study on junior high school teachers. *European Journal of Educational Research*, 9(4), 1491–1502. <https://doi.org/10.12973/eu-jer.9.4.1491>
- Tashtoush, M. A., Shirawia, N., & Rasheed, N. M. (2025). Scoring Rubrics Method in Performance Assessment and its effect of Mathematical Achievement. *Athens Journal of Education*, 12(1), 39–60. <https://doi.org/10.30958/aje.12-1-3>
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615–631. <https://doi.org/10.1080/01411920120095780>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Vlachopoulos, D., & Makri, A. (2024a). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83(August), 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Vlachopoulos, D., & Makri, A. (2024b). Studies in Educational Evaluation A systematic literature review on authentic assessment in higher education : Best practices for the development of 21st century skills , and policy considerations. *Studies in Educational Evaluation*, 83(August), 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Wang, S., Hao, X., Ma, X., Yu, Y., Wu, L., Wang, Y., & Li, Y. (2020). Associations between Poor Vision , Vision-Related Behaviors and Mathematics Achievement in Chinese Students from the CNAEQ-PEH 2015. *International Journal of Environmental Research and Public Health* Article, 17, 1–18. <https://doi.org/10.3390/ijerph17228561>
- Wiewiora, A., & Kowalkiewicz, A. (2019). The role of authentic assessment in developing authentic leadership identity and competencies. *Assessment and Evaluation in Higher Education*, 44(3), 415–430. <https://doi.org/10.1080/02602938.2018.1516730>
- Wu, X., Xu, T., & Wu, R. (2024). Research on personalized assessment Information of students' mathematical competency based on cognitive diagnosis. *Thinking Skills and Creativity*, 53(August), 101596. <https://doi.org/10.1016/j.tsc.2024.101596>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Yudha, R. P., Anggara, D. S., & Zulaeha, O. (2019). Authentic assessment instruments for performance in mathematics learning in elementary schools Authentic assessment instruments for performance mathematics learning in elementary schools. *Journal of Physics: Conference Series*, 1321(032012), 1–13. <https://doi.org/10.1088/1742-6596/1321/3/032012>