

Fostering Mathematical Literacy in Primary Education through Classroom Assessment: A Systematic Literature Review

Retni Paradesa^{1,2}, Zulkardi¹, Ratu Ilma Indra Putri¹, Darmawijoyo¹

¹ Universitas Sriwijaya, Indonesia

² Universitas Islam Negeri Raden Fatah Palembang, Indonesia

ARTICLE INFO

Keywords:

Mathematical literacy;
Classroom Assessment;
Primary school students;
Digital assessment;
Feedback literacy;
Self-regulated learning

Article history:

Received 2025-04-18
Revised 2026-08-22
Accepted 2026-10-21

ABSTRACT

Mathematical literacy requires primary students to formulate, employ, and interpret mathematics in meaningful situations, yet classroom assessment often remains dominated by routine and procedural tasks. This systematic literature review examined how classroom assessment, including technology-enhanced practices, can foster mathematical literacy in primary education. Scopus and ERIC were searched between January and March 2025 for studies published from 2014 to 2025. Following PRISMA 2020 procedures, 1,200 records were identified, 230 full texts were assessed, and 51 studies meeting the eligibility and quality criteria were synthesized thematically. Five interdependent domains emerged: assessment design and validity; technological features; feedback and feedback literacy; student engagement and self-regulated learning; and teacher assessment literacy and systemic conditions. Classroom assessment supported mathematical literacy when authentic, cognitively demanding tasks were combined with actionable feedback, revision, multimodal representations, and instructional decisions. Digital platforms, automated feedback, adaptive support, and learning analytics strengthened these processes, but technology alone was insufficient when assessment remained test-centered or infrastructure and teacher capacity were limited. The review proposes a Pedagogy–Technology Integration Framework linking assessment design, technological affordances, feedback, learner regulation, and mathematical literacy. The findings support assessment reform aligning curriculum, digital infrastructure, teacher professional learning, and formative practices around reasoning, modelling, and contextual problem solving.

This is an open access article under the [CC BY-NC-SA](#) license.



Corresponding Author:

Retni Paradesa

Universitas Islam Negeri Raden Fatah Palembang, Indonesia; retniparadesa_uin@radenfatah.ac.id

1. INTRODUCTION

Mathematical literacy is a central goal of contemporary primary mathematics education, requiring students to formulate mathematical problems, employ mathematical concepts and procedures, and interpret results in meaningful real-world contexts (OECD, 2023). These processes extend beyond procedural proficiency to encompass reasoning, modeling, communication, and critical judgement. Classroom assessment therefore plays an important role not only in documenting achievement but also in eliciting students' mathematical thinking and generating evidence that can guide subsequent teaching and learning.

Formative and learning-oriented assessment research indicates that assessment is most productive when cognitively meaningful tasks are aligned with learning goals, evidence is used instructionally, and feedback enables students to improve subsequent performance (Black & Wiliam, 2018; Shepard, 2019; Wilson, 2018; Zeng et al., 2018). Digital technologies may strengthen these processes through web-based assessment, multimodal representations, automated feedback, adaptive support, and learning analytics (Fjørtoft, 2020; Weigand et al., 2024; Zhu et al., 2020). However, technology does not inherently improve mathematical literacy. Digital assessment may reproduce conventional procedural testing unless technological affordances are pedagogically aligned with authentic tasks, feedback, reflection, and opportunities for mathematical reasoning.

Existing research has often examined formative assessment, digital learning, feedback, self-regulated learning, teacher assessment literacy, and mathematical literacy separately. Consequently, systematic evidence remains limited on how pedagogical assessment design and technological affordances interact through feedback, engagement, and self-regulation to support mathematical literacy in primary education. Addressing this fragmentation is particularly important as education systems increasingly integrate digital assessment with competency-oriented curricula and international mathematical literacy expectations. Accordingly, this review uses PISA mathematical literacy as a cross-national interpretive lens; national initiatives, including Indonesia's *Tes Kemampuan Akademik* (TKA), are treated as contextual applications rather than the organizing framework.

This systematic review therefore examines: (RQ1) What pedagogical and technological features of classroom assessment support mathematical literacy in primary education? (RQ2) Through what feedback, engagement, and self-regulatory processes do these features influence learning? (RQ3) What implementation challenges, theoretical implications, and policy priorities emerge from the evidence? The review contributes an integrated Pedagogy–Technology Integration Framework connecting assessment design, technological affordances, feedback processes, learner regulation, and mathematical literacy outcomes.

2. METHODS

Review design

This study employed a systematic literature review to synthesize empirical evidence on classroom assessment and mathematical literacy in primary education. The review followed PRISMA 2020 procedures for identification, screening, eligibility assessment, and inclusion. Figure 1 summarizes the review stages, while Figure 2 presents the numerical flow of records. Conceptual background sources were distinguished from the 51 included studies, which constituted the primary dataset.

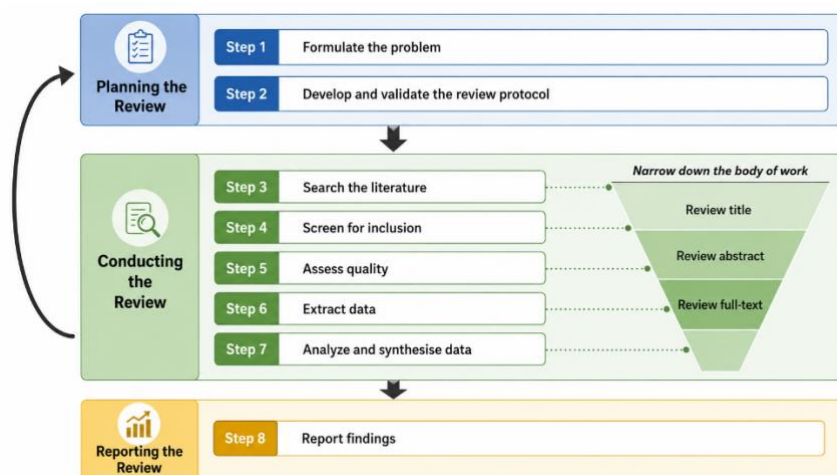


Figure 1. Systematic literature review (SLR) process

Search strategy

Scopus and ERIC were searched between January and March 2025 for publications issued from 2014 to 2025. Search strings combined terms related to classroom and formative assessment, primary mathematics, mathematical literacy, digital or web-based assessment, feedback literacy, self-regulated learning, and technology-enhanced learning. Database-specific syntax was used, and backward and forward citation searching supplemented the electronic search.

The core Boolean structure was:

("classroom assessment" OR "formative assessment" OR "learning-oriented assessment" OR "digital assessment" OR "web-based assessment")

AND (mathematics OR "mathematical literacy" OR numeracy)

AND (primary OR elementary OR K-6)

AND (feedback OR "self-regulated learning" OR reasoning OR modelling OR technology)

Search results were exported, deduplicated, and screened using predefined eligibility criteria. The complete search syntax for each database should be provided in an online supplement to support reproducibility.

Inclusion and exclusion criteria

We included peer-reviewed Scopus- or ERIC-indexed journal articles published in English between 2014 and 2025 that reported empirical findings or validated assessment designs relevant to mathematics assessment, mathematical literacy, or transferable assessment mechanisms for primary education. We excluded studies without relevant assessment components, transferable mechanisms, empirical evidence, or sufficient methodological information.

Study selection process

The search identified 1,200 records. After removing 200 duplicates, we screened 1,000 records, excluded 750, and sought 250 reports. Twenty were unavailable; 230 full texts were assessed, 179 excluded, and 51 studies retained, consistent with Figure 2.

Quality appraisal

We assessed study quality using MMAT, focusing on design, sampling, data collection, analysis, and evidence–conclusion coherence.

Data extraction and synthesis

We extracted data systematically and synthesized them thematically, then mapped them to the PISA processes of formulating, employing, and interpreting.

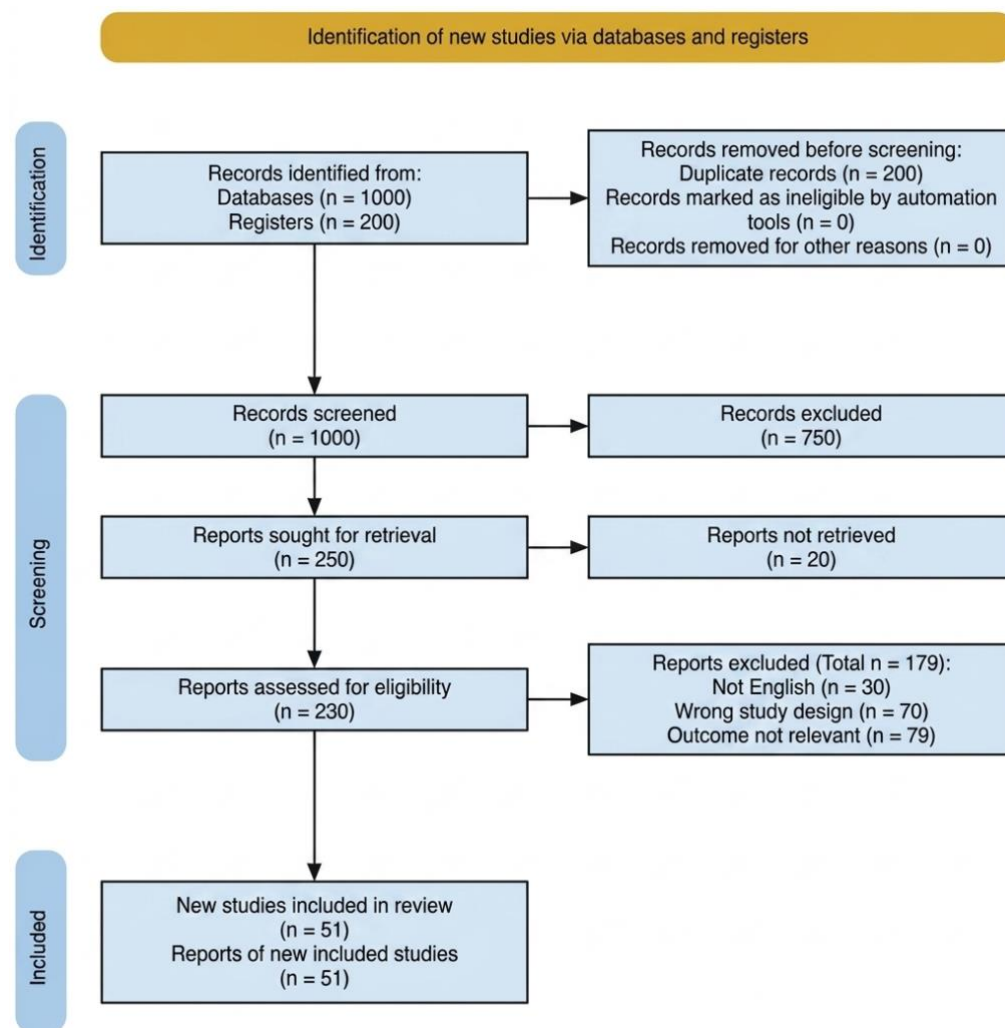


Figure 2. PRISMA 2020 flow diagram.

We analyzed bibliographic data from the included studies in VOS viewer using title and author-keyword terms. We applied full counting, association-strength normalization, and a minimum occurrence threshold of two; a thesaurus standardized synonym, abbreviations, and spelling. The map served as supplementary analysis, while thematic conclusions relied on full-text coding.

3. FINDINGS AND DISCUSSION

The findings are organized into six interconnected sections: characteristics of the included studies, thematic synthesis, technological features supporting mathematical literacy, interpretation through the PISA mathematical literacy processes, the proposed Pedagogy–Technology Integration Framework, and implementation challenges and implications. Across the 51 studies, classroom assessment best supported mathematical literacy when assessment design, feedback, learner participation, and instructional use of evidence aligned coherently. Technology strengthened these processes when it enabled richer representations, timely feedback, revision, adaptation, collaboration, and evidence-

informed teaching, but contributed less when used primarily to reproduce conventional testing practices.

3.1 Characteristics of the Included Studies

The final corpus comprised 51 studies representing qualitative, quantitative, mixed-methods, design-based, assessment-development, and validation approaches. The studies addressed formative assessment, Assessment for Learning (AfL), feedback and feedback literacy, self- and peer-assessment, self-regulated learning (SRL), mathematical reasoning, authentic assessment, technology-enhanced assessment, teacher assessment literacy, learning analytics, and data-informed instruction.

The corpus's diversity enabled identification of recurring assessment mechanisms across educational contexts. However, not all studies focused directly on primary mathematics or explicitly measured mathematical literacy. The synthesis therefore distinguished between studies providing direct evidence on mathematics, primary education, or digital assessment and studies contributing transferable mechanisms related to feedback, self-regulation, assessment literacy, and instructional decision-making. This distinction was necessary to avoid interpreting broader assessment findings as direct evidence of effects on primary students' mathematical literacy.

Several patterns were evident across the studies. Assessment was increasingly conceptualised as an integral component of instruction rather than an activity conducted only after learning. The literature also showed a shift from producing scores to generating and interpreting evidence for subsequent action. Students were given more active roles through reflection, feedback engagement, self-assessment, and revision. At the same time, digital technologies expanded the types of evidence that teachers could collect and use.

Table 1 summarises the methodological characteristics, assessment approaches, reported challenges, and contributions of the 51 included studies.

Table 1. Characteristics and quality assessment (MMAT) of included studies (n=51).

N o	Yea r	Approach	Objective	Challenges	Impact	Citations
1	2018	Assessment for Learning; assessment–pedagogy integration	Linking formative assessment with learning to encourage HOTS and instructional improvement	Misalignment of policy and practice; quality of tasks and feedback	Changing teacher practices to provide feedback that supports learning	(Black & Wiliam, 2018)
2	2018	Feedback literacy	Developing students' capacity to use feedback	Passive feedback use	Active engagement improves learning	(Carless & Boud, 2018)
3	2018	Evaluative judgement	Building students' ability to judge quality	Lack of evaluative skills	Improved self-monitoring	(Tai et al., 2018)
4	2019	Sustainable assessment	Preparing students for lifelong learning	Short-term assessment focus	Long-term learning capacity	(Boud, 2019)

5	2017	Feedback design	Improving feedback effectiveness	Feedback not used	Better learning engagement	(Winstone et al., 2017)
6	2012	Formative assessment theory	Understanding formative processes	Superficial implementation	Stronger learning outcomes	(Clark, 2012)
7	2001	Self-regulated learning	Understanding SRL in learning	Lack of scaffolding	Improved autonomy	(Paris & Paris, 2001)
8	2018	SRL in assessment	Linking SRL with feedback	Weak integration	Better self-monitoring	(Panadero, 2018)
9	2022	Feedback engagement	Improving student response	Low engagement	Active learning	(Hyland; et al., 2022)
10	2017	Peer assessment	Enhancing collaborative learning	Reliability issues	Improved reflection	(Altınay, 2017)
11	2022	Assessment literacy	Defining teacher competencies	Lack of training	Improved practice	(Coombs, A. & DeLuca, 2022)
12	2017	Data use in classrooms	Using data for instruction	Limited data skills	Better decisions	Mandinach & Gummer (2017)
13	2008	Teacher professional development	Improving assessment practices	Time constraints	Enhanced teaching	(Sato et al., 2008)
14	2019	Feedback in classrooms	Enhancing feedback quality	Lack of clarity/actionability	Improved achievement	(Shepard, 2019)
15	2024	Mathematical reasoning	Examining reasoning in classroom tasks	Limited reasoning opportunities	Deeper conceptual understanding	(Stylianides, G. J. & Stylianides, 2024)
16	2007	Problem solving assessment	Assessing mathematical modelling	Procedural dominance	Improved problem-solving skills	(Blum & Leiß, 2007)
17	2023	PISA-oriented assessment	Assessment and mathematical literacy	Task authenticity issues	Better alignment with PISA	(OECD, 2023)
18	2013	Inquiry-based math assessment	Supporting reasoning through inquiry	Teacher readiness	Improved engagement	(Schoenfeld & Kilpatrick, 2013)

19	2023	SDG 4 assessment	Aligning assessment with policy	Equity gaps	Inclusive systems	(Ydesen; & Maren, 2023)
20	2010	Assessment systems	Designing balanced systems	Over-reliance on testing	Better system coherence	(Bennett, 2010)
21	2019	Assessment reform	Improving classroom assessment	Resistance to change	System improvement	(Shepard, 2019)
22	2017	Mathematical thinking	Developing higher-order thinking	Low cognitive demand	Stronger reasoning skills	(Schoenfeld, 2017)
23	2018	Classroom assessment principles to support learning.	Formulating coherent principles linking AfL, grading, and large-scale assessment.	Negative impact of excessive testing; curriculum–assessment coherence.	Guidelines for safe and effective assessment principles and practices.	(Shepard et al., 2018)
24	2018	Classroom-focused educational measurement	Strengthening the role of classroom assessment in instructional decision-making	Ensuring evidence of instruction relevance and validity	Emphasis on the design of instruments that are meaningful for instruction	(Wilson, 2018)
25	2022	Digital assessment	Using tech in assessment	Access issues	Increased engagement	(Davies, 2022)
26	2023	AI in assessment	Automating feedback	Accuracy concerns	Efficiency gains	(Bearman, M; Nieminen, J. H; Ajjawi, 2023)
27	2022	Online formative assessment	Supporting online learning	Interaction limits	Flexible learning	(Veugen, M. J; Gulikers & den Brok, 2022)
28	2016	Framework for the validity of instructional	Conceptualization and evaluation of the validity of assessments relevant to instruction	Multi-source validity evidence; score reasoning–utilization	Operational framework of validity for classroom practice	(Pellegrino et al., 2016)
29	2018	Learning-Oriented	Reviewing LoA components to	Translating LoA into authentic tasks and feedback	Roadmap for effective LoA implementation	(Zeng et al., 2018)

		Assessment (LoA)	enhance learning			
30	2019	Student self-assessment	Assessing evidence of the impact of self-assessment on learning and self-regulation	Self-assessment accuracy; scaffolding needs.	Improved SRL and achievement when supported by appropriate strategies.	(Andrade, 2019)).
31	2018	Assessment literacy instrument (ACAI)	Developing a teacher assessment literacy instrument	Variations in teacher assessment competencies	Diagnostic tools for targeted PD	(DeLuca et al., 2018)).
32	2018	AfL and data-based decision making.	Exploring teachers' views on AfL and data use.	Data analytics capacity; routine integration in the classroom	More targeted data-based practices for teaching	(Kippers et al., 2018)
33	2020	Assessment–learning–self-regulation framework	Conceptually linking classroom assessment with SRL	Task/feedback design that triggers SRL	Pro-SRL assessment design model	(Chen & Bonner, 2020)
34	2018	Fusion of SRL and formative assessment	Providing a roadmap for integrating SRL into AfL	SRL in daily assessment practices.	Map of SRL strategy integration in AfL	(Panadero, 2018).
35	2021	Feedback literacy (evidence-based framework)	Improving student feedback literacy	Low feedback literacy; affective barriers; limited practice.	More productive use of feedback	(Yu & Liu, 2021)
36	2020	Automated formative feedback	Testing the impact of automated feedback on revision and achievement	Automated feedback accuracy varies; integration/infrastructure gaps.	Improved revision behavior and learning gains	(Zhu et al., 2020)
37	2020	Multimodal digital assessment:	Exploring multimodal assessment design in the classroom.	High scoring workload; inter-rater issues; privacy/access concerns.	Enriching assessment evidence and student engagement	(Fjørtoft, 2020)
38	2019	Teachers' assessment mindset	Examining the relationship between mindset and assessment approaches	Uneven teacher assessment literacy; time/PD limits; accountability pressure.	Positive mindset correlates more strongly with AfL practices	(DeLuca et al., 2019)

39	2015	Multilevel model of contextual factors—assessment practices	Summarizing contextual factors that influence teacher practices	Low formative fidelity; belief/norm barriers; weak classroom tools.	Identifying factors that drive AfL implementation	(Fulmer et al., 2015)
40	2020	Culturally responsive assessment	Supporting fair and culturally responsive assessment	Linguistic/cultural bias; inconsistent accommodations; consequential validity risks.	Recommendations for fair and inclusive assessment design	(Nortvedt et al., 2020)
41	2018	Policy expectations & AfL support	Mapping policy expectations & implementation support	Policy–practice gap; workload limits timely, dialogic feedback.	Recommendations for strengthening systemic AfL support	(Kleij et al., 2018)
42	2019	Assessment as Learning (AaL)	Placing students at the center of elementary school writing assessment	Low self/peer reliability; social bias; weak instructional use.	Increased learning engagement and ownership pre-service teachers.	(Lee et al., 2019)
43	2019	Multi-measure teaching effectiveness	Examining various measures of teaching effectiveness	Limited time; behavior-management demands; feedback uptake depends on relationships.	Triangulation strengthens inferences about teaching quality	(Sandilos et al., 2019)
44	2015	Data literacy intervention for prospective teachers	Improving data literacy for instructional decisions	Superficial data use; limited analytic capacity; fragmented systems.	Improved capacity for data-based teaching	(Reeves & Honig, 2015)
45	2007	Formative feedback	Learning support	Weak design	Better outcomes	(Hattie, J. & Timperley, 2007)
46	2016	Classroom dialogue	Feedback interaction	Passive students	Active dialogue	(Simpson, 2016)
47	2019	Assessment culture	School practices	Resistance	Improved culture	(Brown et al., 2019)
48	2017	Learning analytics	Data-driven learning	Data overload	Insights	(Zotou et al., 2020)
49	2020	Adaptive assessment	Personalization	Complexity	Better fit	(Zhai et al., 2021)
50	2021	Competency-based assessment	Skills focus	Implementation gap	Better alignment	(Holmes et al., 2021)

51	2018	Authentic assessment	Real-world tasks	Design complexity	Deeper learning	(Gulikers et al., 2018)
----	------	----------------------	------------------	-------------------	-----------------	-------------------------

We further explored the literature's conceptual structure using VOSviewer. Figure 3 presents the cleaned term co-occurrence network generated after standardising synonymous terms, abbreviations, spelling variations, and duplicate expressions through a thesaurus procedure. The visualisation shows related but partially separated clusters around assessment literacy, feedback literacy, student self-assessment, self-regulated learning, higher-order thinking, and digital assessment.

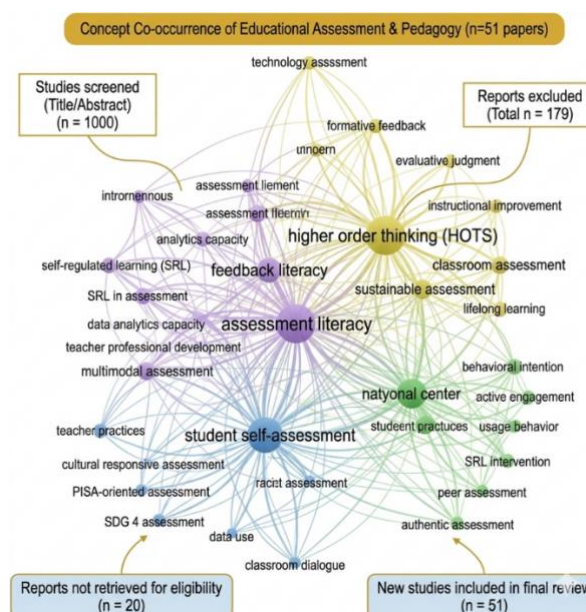


Figure 3. Conceptual structure of classroom assessment research: VOSviewer term co-occurrence network (n = 51 studies).

The network indicates conceptual proximity rather than causal relationships. Its main contribution is to show that research has accumulated around distinct pedagogical and technological domains. At the same time, relatively few studies integrate assessment design, technological affordances, feedback, learner regulation, and mathematical-literacy outcomes in a single explanatory model. This fragmentation informed both the thematic synthesis and the integrated framework presented below.

3.2 Thematic Synthesis

Five interdependent domains emerged from the thematic analysis: (1) assessment design and validity, (2) technological affordances, (3) feedback processes and feedback literacy, (4) student engagement and self-regulated learning, and (5) teacher assessment literacy and systemic conditions. Mathematical literacy was interpreted as the learning outcome generated through interaction among these domains rather than as a separate assessment component.

Assessment design determined the types of mathematical thinking that could become visible. Contextual, non-routine, modelling, constructed-response, and reasoning-oriented tasks produced richer evidence than assessments that emphasized short procedural responses. These tasks enabled students to identify relevant information, represent situations mathematically, select strategies, justify solutions, and interpret results. Accordingly, validity depended not only on technical properties of assessment instruments but also on alignment among learning objectives, task demands, evidence collected, and instructional decisions.

Feedback constituted the mechanism through which assessment evidence could influence subsequent learning. Across the literature, effective feedback was timely, understandable, criteria-referenced, actionable, and connected with opportunities for revision. However, feedback provision alone was insufficient. Students also needed feedback literacy—the capacity to interpret information, judge the quality of their work, and take appropriate action. Rubrics, exemplars, teacher–student dialogue, peer discussion, and opportunities to revise solutions were therefore important components of effective formative assessment.

Student engagement and SRL further mediated the relationship between assessment and learning. Self-assessment, reflection, goal setting, monitoring, repeated attempts, and peer dialogue encouraged students to participate actively in evaluating their learning. These processes are particularly relevant to mathematical literacy because solving contextual problems requires learners to select strategies, monitor progress, evaluate the plausibility of results, and modify approaches when necessary. Nevertheless, self- and peer-assessment required scaffolding because students did not automatically possess sufficient evaluative judgement.

Teacher assessment literacy operated across all domains. Teachers required competence in designing tasks, interpreting evidence, providing feedback, using technologies appropriately, ensuring fairness, and adapting instruction. Broader factors—including professional development, workload, curriculum requirements, school leadership, infrastructure, and assessment culture—determined whether teachers could use assessment evidence for formative purposes. Thus, sustainable improvement in mathematical literacy depends on coherence among classroom practice, professional capacity, technological support, and educational policy.

3.3 Technological Features Supporting Mathematical Literacy

Technology-focused studies indicated a shift from simple online testing toward digitally supported assessment environments. Their educational value depended less on the technology itself than on the pedagogical functions it enabled.

Web-based assessment platforms and learning management systems improved accessibility, task distribution, response collection, scoring, and communication. However, their contribution to mathematical literacy was limited when conventional multiple-choice tests were merely transferred to digital formats. Platforms added value when they supported open responses, representations, explanation, collaboration, feedback, and revision, making students' mathematical thinking more visible.

Adaptive assessment and AI-assisted systems supported personalization by adjusting task difficulty, identifying error patterns, generating prompts, and recommending subsequent activities. Nevertheless, adaptation based only on correctness or response speed did not necessarily strengthen reasoning. For mathematical literacy, adaptive systems need to respond to evidence of students' strategies, representations, and misconceptions.

Automated feedback reduced the time between response and guidance, supporting repeated attempts and correction of procedural or representational errors. Yet automated systems were less suitable for evaluating complex modeling, explanation, and justification. They are therefore better positioned as supplements to teacher and peer judgement than as replacements for professional assessment.

Learning analytics and dashboards supported monitoring of participation, progress, misconceptions, revision, and use of feedback. Their usefulness depended on alignment with learning goals. Risks included data overload, weak construct alignment, and overinterpreting behavioral indicators. Digital traces such as time-on-task should therefore be combined with substantive evidence of reasoning.

Gamification, mobile learning, and collaborative online environments generally increased participation, persistence, and opportunities for mathematical discourse. Their contribution was stronger when activities required strategic choice, explanation, critique, and reflection rather than speed or reward accumulation.

Multimodal assessment and digital portfolios provided rich evidence through written work, diagrams, recordings, images, and revision histories, although they increased demands for rubric quality, moderation, workload, privacy, and accessibility.

Overall, technology contributed most when it improved evidence quality, strengthened feedback and revision cycles, and aligned with the construct being assessed. A major research gap remains in comparative and longitudinal studies examining how specific technological features influence distinct dimensions of mathematical literacy in primary and resource-constrained contexts.

3.4 Mathematical Literacy as the Integrating Outcome

The findings were interpreted through the PISA mathematical literacy processes of formulate, employ, and interpret, providing an internationally relevant lens for understanding the contribution of classroom assessment.

For formulate, assessment tasks should require learners to identify relevant information, make assumptions, construct representations, and translate real situations into mathematical forms. Tasks that provide all mathematical structures in advance offer limited evidence of this process.

For employ, assessment should reveal how students select and apply concepts, procedures, and strategies. Evidence should therefore include solution processes, representations, strategic choices, and justification, not only final answers.

For interpret, students must evaluate mathematical results in relation to the original context, determine whether conclusions are reasonable, and communicate their implications or limitations. Assessment should consequently provide opportunities for contextual explanation and reflection.

Technology can support all three processes through dynamic representations, authentic datasets, interactive modelling, multimodal responses, and rapid feedback. Conversely, highly constrained digital interfaces may reduce opportunities for demonstrating reasoning when they privilege fixed responses and obscure intermediate work.

Table 2. Mapping classroom assessment and technological support to PISA mathematical literacy processes

PISA process	Illustrative assessment task	Evidence of learning	Supporting assessment practice	Potential technological affordance
Formulate	Contextual modelling problem	Relevant information, assumptions, representations	Criteria-referenced rubric and questioning	Dynamic representations, authentic datasets
Employ	Non-routine multi-step problem	Strategy selection, mathematical operations, alternative solutions	Feedback and misconception analysis	Adaptive prompts, automated procedural feedback

Interpret	Data/graph interpretation task	Reasonableness, contextual justification, communication	Reflection and discussion	Multimodal responses, collaborative discussion
------------------	--------------------------------	---	---------------------------	--

The international significance of this mapping extends beyond preparation for any single national assessment. National assessment initiatives, including Indonesia's TKA, may apply these principles as contextual examples, but they do not define this review's analytical framework.

3.5 Pedagogy–Technology Integration Framework

The synthesis generated a Pedagogy–Technology Integration Framework for Fostering Mathematical Literacy in Primary Education, which integrates domains often examined separately and conceptualizes classroom assessment as an interconnected learning system. Its central sequence is:

Assessment Design → Technological Affordances → Feedback Processes → Student Engagement and Self-Regulated Learning → Mathematical Literacy Outcomes

Teacher assessment literacy and systemic conditions operate across the entire sequence.

Assessment design forms the pedagogical foundation because task quality determines the mathematical thinking students are invited to demonstrate. Authentic, contextual, and cognitively demanding tasks create opportunities for modelling, reasoning, and interpretation. Technological affordances extend these opportunities through interactive representations, multimodal responses, learning analytics, automated support, and adaptive systems. However, technology is not treated as an independent cause of learning; its value depends on pedagogical purpose and professional judgement.

Feedback links assessment evidence with subsequent learner action. Criteria, dialogue, prompts, and opportunities for revision help students reconsider strategies and improve their work. These processes support engagement and self-regulated learning by encouraging students to monitor progress, evaluate alternatives, and adjust strategies. Together, they strengthen students' ability to formulate mathematical situations, employ concepts and strategies, and interpret results meaningfully.

Teacher assessment literacy is positioned as a cross-cutting requirement because teachers select tasks and technologies, interpret evidence, design feedback, and determine instructional responses. Systemic factors, including professional development, curriculum alignment, infrastructure, policy, assessment culture, and equity, influence sustainable implementation. The framework's theoretical contribution lies in explaining how pedagogy and technology operate through feedback and learner regulation to foster mathematical literacy.

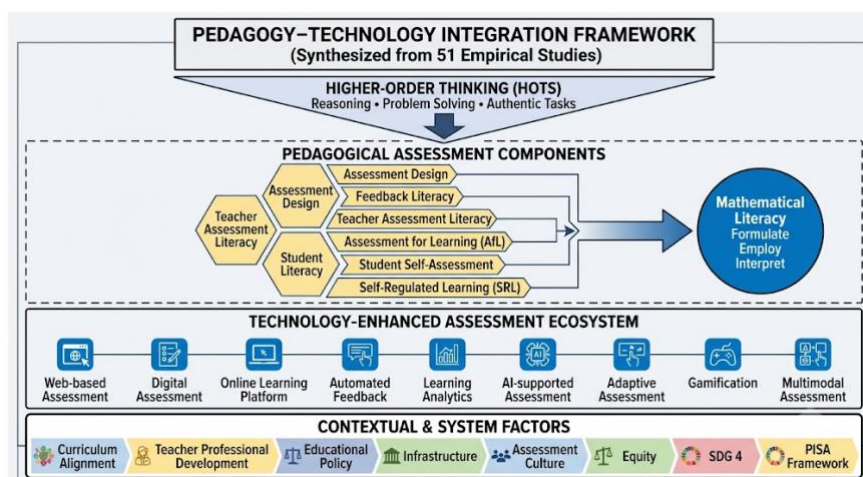


Figure 4. Pedagogy–Technology Integration Framework for Fostering Mathematical Literacy in Primary Education.

The framework's theoretical contribution lies in explaining how pedagogy and technology operate through feedback and learner regulation rather than treating assessment design, technology, feedback, SRL, and mathematical literacy as parallel constructs.

3.6 Implementation Challenges and Evidence-Based Implications

Four recurrent implementation challenges emerged: teacher capacity, technological infrastructure, validity and fairness, and systemic coherence.

Teacher capacity was particularly important. Effective technology-enhanced assessment requires not merely operational knowledge of digital platforms but digital assessment literacy, including task design, feedback, evidence interpretation, data literacy, fairness, and awareness of technological limitations. Professional development should therefore emphasize pedagogically informed use rather than technical training alone.

Infrastructure and access represented a second challenge. Reliable connectivity, appropriate devices, accessible platforms, and technical support are prerequisites for equitable digital assessment. Without these conditions, digital innovation may reinforce existing inequalities.

Validity and fairness also require attention. Automated scores and behavioral analytics should not be assumed to represent mathematical understanding directly. Complex competencies require multiple forms of evidence, including constructed responses, student explanations, observations, artefacts, and digital traces.

Finally, systemic coherence is necessary. Curriculum, formative assessment, summative assessment, teacher education, professional development, infrastructure, and policy need to reinforce rather than contradict one another. Systems that emphasize score reporting without supporting feedback and revision constrain learning-oriented assessment.

The findings consequently imply several policy priorities. Governments and education authorities should invest in equitable digital infrastructure and reliable connectivity. Curriculum developers should integrate reasoning, modelling, problem-solving, communication, collaboration, critical thinking, and digital competence into primary mathematics assessment. Teacher education should strengthen assessment literacy, feedback literacy, data interpretation, digital pedagogy, and responsible use of AI-supported assessment. School leaders should provide time and structures for collaborative task design, moderation, evidence interpretation, and professional reflection.

At the international level, PISA offers a useful reference because mathematical literacy emphasizes using mathematics meaningfully rather than merely reproducing procedures. National initiatives can adapt these principles to their particular contexts. In Indonesia, TKA represents one possible application of this broader approach rather than the conceptual foundation of the review.

Overall, the synthesis indicates that mathematical literacy develops most effectively within an assessment ecology in which pedagogy establishes purpose, technology expands assessment possibilities, feedback connects evidence with action, learners actively regulate their learning, teachers exercise informed professional judgement, and educational systems provide the conditions required for equitable and sustainable implementation.

Discussion

This systematic review shows that classroom assessment fosters mathematical literacy when it functions as an integrated learning-oriented system rather than merely measuring achievement. Effective practices combine cognitively demanding contextual tasks, formative feedback, reflection, self-assessment, and evidence-informed instructional decisions, enabling students to formulate, employ, and interpret mathematics in meaningful situations (OECD, 2019, 2023). These findings suggest that mathematical literacy develops through the interaction of assessment design, feedback, student agency, and instructional response (Black & Wiliam, 2018; Shepard, 2019).

Technology serves as an enabling layer within this system. Web-based assessment, automated feedback, multimodal assessment, learning analytics, and adaptive tools can enrich evidence collection, accelerate feedback, and support revision, but their value depends on pedagogical alignment with authentic mathematical tasks (Fjørtoft, 2020; Zhu et al., 2020).

Feedback and self-regulated learning mediate how students use assessment evidence. Productive feedback requires students to interpret information and revise their strategies, while teachers need strong assessment literacy to design tasks, interpret evidence, and support improvement (Carless, D. & Boud, 2018).

Accordingly, the proposed Pedagogy–Technology Integration Framework links assessment design, technological affordances, feedback, self-regulation, and mathematical literacy, with teacher capacity and institutional conditions operating across the system. It also highlights the need for equitable infrastructure and professional learning.

4. CONCLUSION

This review shows that classroom assessment can foster mathematical literacy when authentic tasks, formative feedback, reflection, self-regulation, and instructional decisions are coherently integrated. When pedagogically aligned, technology strengthens these processes through interactive assessment, timely feedback, multimodal evidence, and learning analytics. The proposed Pedagogy–Technology Integration Framework explains how assessment design, technological affordances, feedback, engagement, and self-regulated learning interact within supportive teacher and institutional conditions. Limitations include the restriction to English-language studies from Scopus and ERIC and variation across research contexts. Future research should empirically test the framework and examine equitable implementation across diverse primary education settings.

Acknowledgments: The authors gratefully acknowledge the financial support provided by *Kementerian Agama Republik Indonesia* through *Program Bantuan Penyelesaian Pendidikan S3 Dalam Negeri (BPP S3 Dalam Negeri)*. The authors also express sincere appreciation to colleagues and academic mentors for their valuable feedback and support during the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

- Altınay, Z. (2017). Evaluating peer learning and assessment in online collaborative learning environments. *Behaviour & Information Technology*, 36(3), 312–320. <https://doi.org/https://doi.org/10.1080/0144929X.2016.1232752>
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. In *Frontiers in Education* (Vol. 4). <https://doi.org/10.3389/feduc.2019.00087>
- Bearman, M; Nieminen, J. H; Ajjawi, R. (2023). Designing assessment in a digital world: an organising framework. *Assessment & Evaluation in Higher Education*, 48(1), 291–304. <https://doi.org/https://doi.org/10.1080/02602938.2022.2069674>
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8, 70–91.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Blum, W., & Leiß, D. (2007). *How do Students and Teachers Deal with Modelling Problems?* (C. Haines, P. Galbraith, W. Blum, & S. B. T.-M. M. Khan (eds.); pp. 222–231). Woodhead Publishing. <https://doi.org/https://doi.org/10.1533/9780857099419.5.221>
- Boud, D. (2019). Sustainable assessment: Revisiting assessment for the learning society. *Assessment & Evaluation in Higher Education*, 44(5), 761–764. <https://doi.org/https://doi.org/10.1080/02602938.2019.1610164>
- Brown, G. T. L., Gebril, A., & Michaelides, M. P. (2019). Teachers' Conceptions of Assessment: A Global Phenomenon or a Global Localism. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00016>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Chen, P. P., & Bonner, S. M. (2020). A framework for classroom assessment, learning, and self-regulation. *Assessment in Education: Principles, Policy and Practice*, 27(4), 373–393. <https://doi.org/10.1080/0969594X.2019.1619515>
- Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educ Psychol Rev*, 24, 205–249. <https://doi.org/https://doi.org/10.1007/s10648-011-9191-6>
- Coombs, A., & DeLuca, C. (2022). Mapping the constellation of assessment discourses: a scoping review study on assessment competence, literacy, capability, and identity. *Educ Asse Eval Acc*, 34, 279–301. <https://doi.org/https://doi.org/10.1007/s11092-022-09389-9>
- Davies, A. L. T. B. J. von. (2022). Digital-First Learning and Assessment Systems for the 21st Century. *Front. Educ.*, 7, 7:857604. <https://doi.org/10.3389/feduc.2022.857604>
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy and Practice*, 25(4), 355–375. <https://doi.org/10.1080/0969594X.2016.1244514>
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, Programs, and Practices: Exploring the Complex Dynamics of Assessment Education in Teacher Education Across Four Countries. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00132>
- Fjørtoft, H. (2020). Multimodal digital classroom assessments. *Computers and Education*, 152. <https://doi.org/10.1016/j.compedu.2020.103892>
- Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: an integrative review of research. *Assessment in Education: Principles, Policy and Practice*, 22(4), 475–494. <https://doi.org/10.1080/0969594X.2015.1017445>

- Gulikers, J. T. M., Runhaar, P., & Mulder, M. (2018). An assessment innovation as flywheel for changing teaching and learning. *Journal of Vocational Education & Training*, 70(2), 212–231. <https://doi.org/https://doi.org/10.1080/13636820.2017.1394353>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Holmes, A. G. D., Polman Tuin, M., & Turner, S. L. (2021). Competence and competency in higher education, simple terms yet with complex meanings: Theoretical and practical issues for university teachers and assessors implementing Competency-Based Education (CBE). *Educational Process: International Journal*, 10(3), 39–52. <https://doi.org/https://doi.org/10.1007/s10956-020-09879-9>
- Hyland, Zhe (Victor) Zhang, & Ken. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51. <https://doi.org/https://doi.org/10.1016/j.asw.2021.100586>
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>
- Kleij, F. M. Van der, Cumming, J. J., & Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assessment in Education Principles Policy and Practice*, 25(6), 620–637. <https://doi.org/10.1080/0969594X.2017.1374924>
- Lee, I., Mak, P., & Yuan, R. E. (2019). Assessment as learning in primary writing classrooms: An exploratory study. *Studies in Educational Evaluation*, 62, 72–81. <https://doi.org/10.1016/j.stueduc.2019.04.012>
- Nortvedt, G. A., Wiese, E., Brown, M., Burns, D., McNamara, G., O'Hara, J., Altrichter, H., Fellner, M., Herzog-Punzenberger, B., Nayir, F., & Taneri, P. O. (2020). Aiding culturally responsive assessment in schools in a globalising world. *Educational Assessment, Evaluation and Accountability*, 32(1), 5–27. <https://doi.org/10.1007/s11092-020-09316-w>
- OECD. (2019). PISA 2018 Assessment and Analytical Framework. In OECD Publishing.
- OECD. (2023). *PISA 2022 assessment and analytical framework: Mathematics, reading, science, financial literacy and creative thinking*. OECD Publishing. <https://doi.org/https://doi.org/10.1787/dfe0bf9c-en>
- Panadero, E. (2018). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422. <https://doi.org/https://doi.org/10.3389/fpsyg.2017.00422>
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36(2), 89–101. https://doi.org/https://doi.org/10.1207/S15326985EP3602_4
- 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>
- Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, 50, 90–101. <https://doi.org/10.1016/j.tate.2015.05.007>
- Sandilos, L. E., Sims, W. A., Norwalk, K. E., & Reddy, L. A. (2019). Converging on quality: Examining multiple measures of teaching effectiveness. *Journal of School Psychology*, 74, 10–28. <https://doi.org/10.1016/j.jsp.2019.05.004>
- Sato, M., Ruth Chung Wei, & Linda Darling-Hammond. (2008). Improving Teachers' Assessment Practices Through Professional Development: The Case of National Board Certification. *American Educational Research Journal*, 45(3), 669–700. <https://doi.org/https://doi.org/10.3102/0002831208316955>
- Schoenfeld, A. H. (2017). Uses of video in understanding and improving mathematical thinking and teaching. *J Math Teacher Educ*, 20, 415–432. <https://doi.org/https://doi.org/10.1007/s10857-017-9381-3>
- Schoenfeld, A. H., & Kilpatrick, J. A. (2013). US perspective on the implementation of inquiry-based learning in mathematics. *ZDM Mathematics Education*, 45, 901–909. <https://doi.org/https://doi.org/10.1007/s11858-013-0531-5>

- Shepard, L. A. (2019). Classroom Assessment to Support Teaching and Learning. *Annals of the American Academy of Political and Social Science*, 683(1), 183–200. <https://doi.org/10.1177/0002716219843818>
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>
- Simpson, A. (2016). Dialogic teaching in the initial teacher education classroom: “Everyone’s Voice will be Heard.” *Research Papers in Education*, 31(1), 89–106. <https://doi.org/https://doi.org/10.1080/02671522.2016.1106697>
- Stylianides, G. J., & Stylianides, A. J. (2024). Promoting elements of mathematical knowledge for teaching related to the notion of assumptions. *Mathematical Thinking and Learning*, 26(4), 382–410.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero. (2018). E Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76, 467–481. <https://doi.org/https://doi.org/10.1007/s10734-017-0220-3>
- Veugen, M. J; Gulikers, J. T. M., & den Brok, P. (2022). Secondary school teachers’ use of online formative assessment during COVID-19 lockdown: Experiences and lessons learned. *Journal of Computer Assisted Learning*, 38(5), 1465–1481. <https://doi.org/https://doi.org/10.1111/jcal.12699>
- Weigand, H. G., Trgalova, J., & Tabach, M. (2024). Mathematics teaching, learning, and assessment in the digital age. *ZDM - Mathematics Education*, 56(4), 525–541. <https://doi.org/10.1007/s11858-024-01612-9>
- Wilson, M. (2018). Making Measurement Important for Education: The Crucial Role of Classroom Assessment. *Educational Measurement: Issues and Practice*, 37(1), 5–20. <https://doi.org/10.1111/emip.12188>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners’ agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/https://doi.org/10.1080/00461520.2016.1207538>
- Ydesen, C., & Maren, E. (2023). SDG 4 as a global governance tool and the quest for recognizing diversity – Implications emerging from the intersections between inclusive education and assessment. *International Journal of Educational Development*, 103. <https://doi.org/https://doi.org/10.1016/j.ijedudev.2023.102932>
- Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48. <https://doi.org/10.1016/j.asw.2021.100525>
- Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students’ learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, 30(3), 211–250. <https://doi.org/10.1007/s11092-018-9281-9>
- Zhai, X., Krajcik, J., & Pellegrino, J. W. (2021). On the Validity of Machine Learning-based Next Generation Science Assessments: A Validity Inferential Network. *J Sci Educ Technol*, 30, 298–312. <https://doi.org/https://doi.org/10.1007/s10956-020-09879-9>
- Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers and Education*, 143. <https://doi.org/10.1016/j.compedu.2019.103668>
- Zotou, M., Tambouris, E., & Tarabanis, K. (2020). Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev*, 68, 3393–3424. <https://doi.org/https://doi.org/10.1007/s11423-020-09828-8>