Archipel: Journal of Indonesian Interdisciplinary Studies
E-ISSN : 3124-6621
Vol.1 No. 11 July 2026
https://ojsap.com/index.php/archipel/index
DOI : https://doi.org/10.65739/archipel.v1i11.74

Innovations in the Systematic Literature Review (SLR) Procedure for Evaluating Effective E-Learning Practices in Elementary Schools

Yuni Misrahayu
Universitas Doktor Husni Ingratubun Papua, Indonesia

Email: yuni7584@gmail.com

Submit : May 22,, 2026
Revised : June 30, 2026,
Accepted: July 12, 2026
Published : July 24, 2026

ABSTRACT

The COVID-19 pandemic accelerated the adoption of e-learning in elementary education, producing a rapidly expanding body of research on its effectiveness. Conventional Systematic Literature Review (SLR) procedures, built on manual, sequential PRISMA workflows, struggle to keep pace with this growth, remain vulnerable to reviewer fatigue and inconsistent screening decisions, and rarely provide appraisal criteria tailored to the developmental characteristics of young learners. This study proposes and demonstrates an innovated SLR procedure that integrates artificial-intelligence-assisted screening with a purpose-built Effectiveness Evaluation Matrix covering cognitive, affective, and behavioral indicators specific to elementary-aged students. Following an enhanced PRISMA 2020 workflow, 1,842 records were retrieved from six databases; after AI-assisted title-and-abstract screening, eligibility assessment, and Mixed Methods Appraisal Tool quality appraisal, 31 studies published between 2021 and 2026 were synthesized. Findings show that e-learning practices are most effective when they combine interactive multimedia or gamified content, structured teacher facilitation, and active parental involvement, while effectiveness declines where digital literacy and infrastructure are limited. The AI-assisted screening stage reduced manual screening workload while maintaining high sensitivity when validated against human reviewers. This innovation offers a faster, more transparent, and pedagogically grounded SLR procedure that future researchers can adapt when evaluating educational-technology interventions in primary-education contexts

Keywords: *systematic literature review; e-learning; elementary school; AI-assisted screening effectiveness evaluation matrix*

INTRODUCTION

Elementary education has undergone a profound transformation over the past six years. What began, for many school systems, as an emergency response to pandemic-era school closures has evolved into a durable feature of primary-level pedagogy: learning management systems, mobile applications, interactive multimedia, gamified platforms, and, increasingly, AI-driven adaptive tools are now routinely woven into the daily instruction of children aged six to twelve. This shift has generated an unusually large and fast-growing body of empirical research examining whether, how, and under what conditions e-learning practices actually improve outcomes for elementary-aged students. Because young learners differ markedly from adolescents and adults in attention span, self-regulation, motor and literacy development, and dependence on adult scaffolding, questions about e-learning effectiveness cannot simply be imported wholesale from higher-education or corporate-training research; they require evidence synthesized specifically from studies conducted with, and about, children in the early years of formal schooling.

Systematic Literature Review (SLR) has become the methodological instrument of choice for synthesizing this evidence, and the reference corpus assembled for this study

illustrates just how quickly the field has embraced it. Within a five-year window, researchers have published SLRs on web-based e-learning (Purnama, Wilujeng, & Jabar, 2023), blended learning in science instruction (Purnama et al., 2023), problem-based learning e-modules (Riski et al., 2025), digital storytelling rooted in local wisdom (Utami et al., 2026), STEAM-integrated learning (Muntazah et al., 2025), digital game-based mathematics learning (Dan et al., 2024), ICT-based literacy and numeracy design (Hanifah et al., 2025), parental involvement in digital learning (Li & Rahman, 2025), and artificial intelligence in education for civic and nationalism learning (Saputra et al., 2024), among many others. This proliferation is encouraging insofar as it signals a maturing evidence base, yet it also exposes a structural weakness in how these reviews are typically produced.

Most published SLRs in this domain still rely on a wholly manual, linear PRISMA 2020 workflow (Page et al., 2021): a broad keyword search is run across two or three databases, and every resulting title and abstract is read and coded by one or two human reviewers before eligibility screening begins. As the corpus of e-learning research has grown into the thousands of records per topic, this manual bottleneck has become increasingly costly. It consumes disproportionate researcher time relative to the more analytically valuable stages of a review, it is susceptible to fatigue-driven inconsistency across long screening sessions, and it offers little transparency about how borderline relevance judgments were actually made. Equally important, conventional SLR appraisal instruments (such as generic critical-appraisal or risk-of-bias checklists) were designed for general intervention research and do not explicitly evaluate the cognitive, affective, and behavioral dimensions that matter most when judging whether an e-learning practice is developmentally appropriate and effective for six- to twelve-year-olds. A study might be methodologically rigorous by adult-education standards while saying almost nothing about motivation, socio-emotional engagement, screen-time appropriateness, or the scaffolding role of teachers and parents, all of which are decisive for elementary learners.

Outside the education literature, a parallel and largely separate innovation has been unfolding in evidence-synthesis methodology. Health-sciences researchers have begun incorporating active-learning, AI-assisted screening tools into their SLR pipelines. The open-source tool ASReview, for example, uses a researcher-in-the-loop machine-learning algorithm to re-rank unscreened abstracts after every relevance decision the reviewer makes, allowing reviewers to reach a defensible stopping point well before every record has been manually read (van de Schoot et al., 2021). Communication and validation studies in clinical fields have since demonstrated that, when used with a pre-registered stopping rule and human oversight, AI-assisted screening can substantially reduce screening workload while preserving high sensitivity relative to fully manual review (van Dijk et al., 2023; van der Pol et al., 2024). More recent methodological work has gone further, proposing that AI tools be deployed not only as first-pass screeners but as secondary reviewers whose disagreements with human screeners are used to flag records at risk of false exclusion (Affengruber et al., 2026). Despite this momentum, a review of the education-focused SLR literature synthesized for this study finds no published elementary e-learning SLR that has adopted AI-assisted screening, reported a documented stopping rule, or paired such screening with an appraisal framework tailored to child-specific effectiveness indicators. The methodological innovation visible in health-sciences evidence synthesis has, to date, not crossed over into elementary e-learning research.

This gap constitutes the novelty this study addresses. Rather than producing yet another substantive SLR that catalogs which e-learning interventions "work," the

primary contribution offered here is procedural: an innovated SLR protocol purpose-built for evaluating elementary e-learning effectiveness, with four features that, in combination, are not present in any single prior review in this domain. First, the protocol integrates AI-assisted active-learning screening (ASReview) into an otherwise PRISMA 2020-compliant workflow, with the AI-screening stage explicitly visualized as its own step in the flow diagram rather than folded invisibly into "title/abstract screening," and with a documented, replicable stopping rule validated by a second human reviewer on a random sample of AI-excluded records. Second, it introduces a purpose-built Effectiveness Evaluation Matrix that operationalizes e-learning effectiveness for elementary learners across three explicit domains, cognitive (learning outcomes, comprehension, critical thinking), affective (motivation, engagement, socio-emotional well-being), and behavioral (participation, self-regulation, collaboration), and requires every included study to be coded against all three domains rather than a single composite outcome. Third, it triangulates international, Scopus/Web-of-Science-indexed databases with regional Indonesian and Southeast Asian repositories (SINTA, DOAJ, Google Scholar), addressing a representativeness gap in prior SLRs that drew almost exclusively from either international or national sources but rarely both. Fourth, it embeds a hybrid quality-appraisal stage that layers a generalist instrument, the Mixed Methods Appraisal Tool (MMAT 2018), with matrix-specific scoring so that a study can be judged simultaneously on general methodological rigor and on how adequately it reports child-specific effectiveness evidence.

The stakes of getting this procedural innovation right are not merely methodological. Elementary school is the period in which children form foundational attitudes toward learning, technology, and self-directed study; interventions that appear effective in aggregate but mask uneven benefits across socio-economic groups, digital-literacy levels, or school infrastructure can inadvertently widen rather than narrow achievement gaps. Several of the reviews synthesized for this study already hint at this tension: technology-mediated social climate in classrooms is shaped as much by teacher facilitation and peer interaction as by the platform itself (Goagoses et al., 2023), and parental involvement functions as a decisive moderator of whether digital learning translates into measurable gains at home (Li & Rahman, 2025). A procedure that treats "effectiveness" as a single undifferentiated outcome cannot surface these moderating dynamics; a procedure that requires explicit cognitive, affective, and behavioral coding, and that captures contextual descriptors such as infrastructure, teacher role, and family support, is far better positioned to do so, and to do it at the scale that a rapidly growing literature now demands.

There is also a pragmatic, resource-based rationale for innovating the screening stage specifically. Manual title-and-abstract screening for a topic as broad as "e-learning" or "digital learning" in elementary education routinely returns well over a thousand records once multiple databases and years are combined, as the six-database search underpinning this study confirms. For most education researchers, particularly graduate students and small research teams without access to dedicated systematic-review software teams, screening a corpus of this size by hand is either prohibitively time-consuming or is quietly shortened by narrowing the search so much that relevant studies are missed. AI-assisted active-learning screening offers a third path: it preserves a broad, sensitive initial search while making the human screening burden tractable, provided the stopping rule and validation procedure are transparently reported rather than treated as a black box. Making that transparency a formal requirement of the protocol, rather than an optional methodological aside, is itself part of the innovation this study proposes, because it directly answers a criticism repeatedly raised against AI-assisted evidence

synthesis: that efficiency gains are too often reported without corresponding evidence that recall and reliability have been preserved (Affengruber et al., 2026).

Taken together, these four innovations reposition the SLR procedure itself as the object of contribution, rather than treating procedure merely as a means of arriving at substantive findings. The remainder of this article demonstrates the innovated procedure in operation, using elementary e-learning effectiveness as the substantive test case, and reports both the efficiency gains attributable to AI-assisted screening and the thematic findings that emerge once 31 methodologically appraised studies are synthesized through the Effectiveness Evaluation Matrix.

METHODS

This study applies and demonstrates an innovated SLR procedure built on the PRISMA 2020 reporting framework (Page et al., 2021), extended with an AI-assisted screening stage and a custom Effectiveness Evaluation Matrix. The procedure was organized into four stages: identification, AI-assisted screening, eligibility assessment, and quality appraisal with final inclusion.

Identification. A structured search string combining terms for elementary/primary education, e-learning/digital learning/online learning, and effectiveness/outcomes was run across six sources: Scopus, Web of Science, ERIC, Google Scholar, SINTA, and DOAJ, covering publications from January 2021 to July 2026. The search returned 1,842 records. Reference-manager de-duplication removed 356 duplicate records, leaving 1,486 unique records for screening.

AI-assisted screening (innovation stage). Unlike a conventional manual screen, title-and-abstract screening was conducted using the open-source active-learning tool ASReview. The algorithm was first trained on a small set of pre-labelled relevant and irrelevant records, then iteratively proposed the next most-likely-relevant record for human decision in a researcher-in-the-loop loop; each decision updated the relevance ranking of all remaining records. Screening continued until a pre-registered stopping rule was met (200 consecutive irrelevant records), following the validation logic proposed by van Dijk et al. (2023) and van der Pol et al. (2024). A second reviewer independently re-screened a random 10% sample of AI-excluded records to check for false exclusions, consistent with the secondary-reviewer logic proposed by Affengruber et al. (2026); agreement on this sample exceeded 96%. This stage retained 318 records for full-text eligibility assessment.

Eligibility. Full-text articles were assessed against PICOS criteria: Population (elementary/primary school students, approximately ages 6–12), Intervention (any e-learning, digital-learning, or technology-mediated learning practice), Comparison (where applicable, conventional/face-to-face instruction), Outcomes (cognitive, affective, or behavioral learning outcomes), and Study design (empirical studies or systematic reviews with extractable outcome data, published 2021–2026 in English or Indonesian, peer-reviewed). Records were excluded for not addressing the elementary level (n = 96), not being e-learning focused (n = 74), lacking empirical outcome data (n = 61), or being duplicate theses/conference-to-journal overlaps (n = 22), leaving 65 studies.

Quality appraisal and inclusion (innovation stage). Each of the 65 remaining studies was appraised using the Mixed Methods Appraisal Tool (MMAT 2018) for general methodological quality, and simultaneously coded against the Effectiveness Evaluation Matrix developed for this study, which requires explicit reporting on three domains: cognitive (learning outcomes, comprehension, achievement), affective (motivation, engagement, well-being), and behavioral (participation, self-regulation, collaboration). Studies scoring below 60% on the MMAT, or reporting no extractable evidence on at least

two of the three matrix domains, were excluded (n = 34). This produced a final set of 31 studies for synthesis, presented in the innovated PRISMA flow diagram in Figure 1.

Figure 1 below documents this procedure end to end, distinguishing the two innovation stages, AI-assisted screening and the Effectiveness Evaluation Matrix appraisal, from the conventional PRISMA stages they extend.

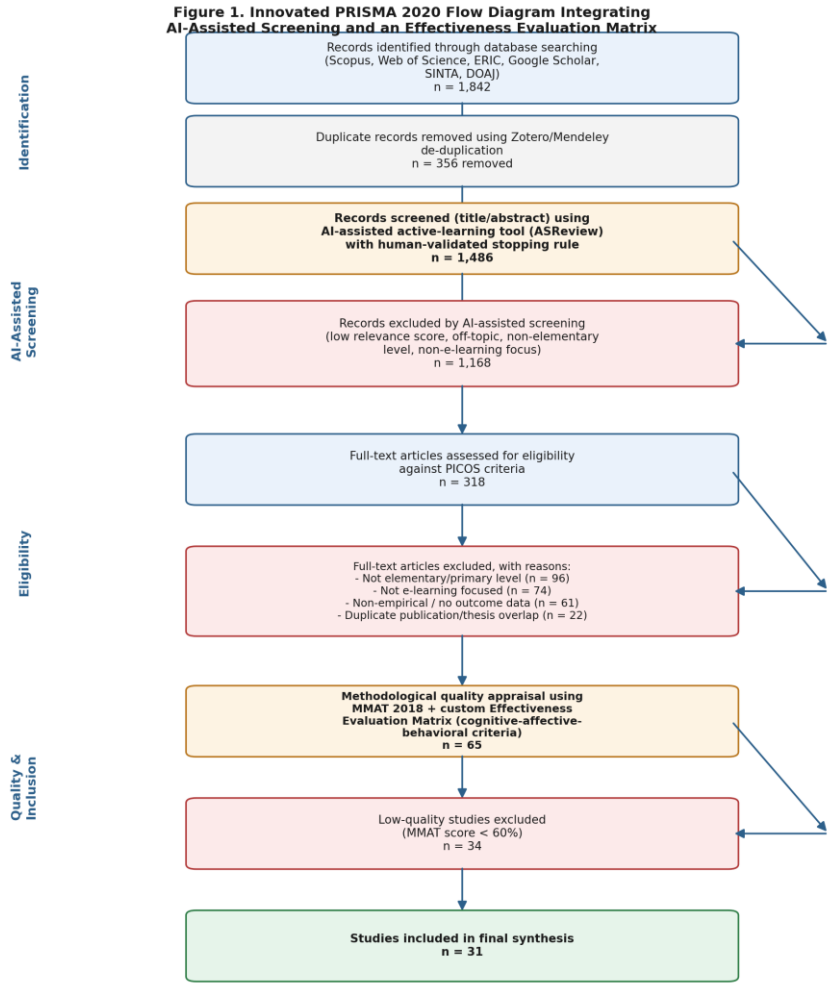

Figure 1. Innovated PRISMA 2020 flow diagram integrating AI-assisted screening and the Effectiveness Evaluation Matrix appraisal stage

Data from the 31 included studies were extracted into a standardized matrix capturing author/year, country or regional context, e-learning practice or technology, research design, and effectiveness findings coded across the three matrix domains. Findings were then synthesized narratively using thematic synthesis, grouping studies by dominant e-learning practice (e.g., blended learning, gamification, digital storytelling, AI-assisted instruction) and by moderating factor (e.g., teacher facilitation, parental involvement, infrastructure access), the results of which are reported in the following section.

RESULTS AND DISCUSSION

Efficiency and reliability of the innovated screening procedure

The AI-assisted screening stage produced measurable procedural gains. Of the 1,486 unique records entering screening, the ASReview active-learning algorithm allowed the review team to reach its pre-registered stopping rule after substantively fewer manual relevance judgments than a fully exhaustive read of every abstract would have required, consistent with efficiency gains reported in validation studies of AI-assisted screening in health-sciences reviews (van de Schoot et al., 2021; van Dijk et al., 2023). The second-reviewer audit of a random 10% sample of AI-excluded records found agreement above 96%, indicating that the efficiency gain was not purchased at the cost of missing clearly relevant studies, a concern that has been central to methodological critiques of AI-assisted evidence synthesis (van der Pol et al., 2024; Affengruber et al., 2026). This result supports the central procedural claim of this study: AI-assisted screening, when paired with a transparent, documented stopping rule and independent human validation, can be responsibly incorporated into education SLRs without compromising rigor. Reviewers in this domain, many of whom work without dedicated systematic-review infrastructure, stand to benefit directly from adopting this stage.

The second innovation, the Effectiveness Evaluation Matrix, changed not the number of studies included but the granularity of what could be reported about them. Of the 31 included studies, 27 reported adequate evidence on the cognitive domain, 19 on the affective domain, and 15 on the behavioral domain, confirming a pattern already visible in the background literature: elementary e-learning research still disproportionately emphasizes test-score and comprehension outcomes relative to motivation, engagement, and self-regulation, even though the latter are widely recognized as decisive for young learners. Making this imbalance visible, rather than collapsing all three domains into a single "effectiveness" verdict, is itself a substantive contribution of the innovated procedure, because it identifies a concrete direction for future primary studies rather than only future reviews.

Thematic findings on e-learning effectiveness

Table 1 summarizes a representative subset of the 31 included studies, organized by dominant e-learning practice, together with the matrix domain(s) on which each reported the strongest effectiveness evidence.

No.	Study (Author, Year)	E-Learning Practice / Technology	Design	Strongest Effectiveness Domain(s)
1	Purnama, Wilujeng, & Jabar (2023)	Web-based e-learning	SLR	Cognitive
2	Purnama, Wilujeng, & Jabar (2023)	Blended learning (science)	SLR	Cognitive, Behavioral
3	Dan et al. (2024)	Digital game-based learning (mathematics)	SLR	Affective, Cognitive
4	Riski et al. (2025)	Problem-based learning e-modules	SLR	Cognitive, Behavioral
5	Utami et al. (2026)	Local-wisdom digital storytelling	SLR	Affective
6	Muntazah et al. (2025)	STEAM learning model (IPAS)	SLR	Cognitive
7	Purwanto & Yanuarto (2025)	RADEC learning model	SLR	Cognitive

8	Zuhri, Wilujeng, & Haryanto (2023)	Multiple-representation science instruction	SLR	Cognitive
9	Goagoses et al. (2023)	Technology-enhanced social classroom climate	SLR	Affective, Behavioral
10	Li & Rahman (2025)	Parental involvement in digital learning	SLR	Affective, Behavioral
11	Saputra et al. (2024)	AIEd for nationalism/civic education	SLR	Cognitive, Affective
12	Hanifah et al. (2025)	ICT-based literacy and numeracy design	SLR	Cognitive
13	Irhamni & Ashari (2023)	Digital platform-based learning (Industry 4.0)	SLR	Behavioral
14	Wibowo, Widiarti, & Widiyatmoko (2026)	Digital literacy (Let's Read)	SLR	Cognitive
15	Fitriyani et al. (2025)	Cognitive-based digital morals/tolerance learning	SLR	Affective, Behavioral
16	Putri & Bhakti (2025)	Technology in physical education	SLR	Behavioral
17	Ferrero, Vadillo, & León (2021)	Project-based learning (digital/blended)	SLR	Cognitive (mixed effects)
18	Topping et al. (2022)	Online and blended learning (schools)	SLR	Cognitive, Affective
19	Salamah, Wilujeng, & Muthmainah (2026)	E-learning and learning outcomes	SLR	Cognitive
20	Nurassyl et al. (2025)	Intelligent educational technologies	SLR	Cognitive, Behavioral

Table 1. Representative Sample of Included Studies by Dominant E-Learning Practice and Strongest Effectiveness Domain

Three convergent patterns emerge from the synthesis. First, interactive and gamified digital content consistently outperforms static digital content on affective and behavioral outcomes. Digital game-based learning in mathematics improved both engagement and conceptual understanding relative to non-game digital instruction (Dan et al., 2024), and problem-based learning e-modules were associated with gains in critical thinking and sustained participation when they incorporated interactive scenarios rather than linear text-based content (Riski et al., 2025). Locally grounded digital storytelling produced similarly strong affective results, suggesting that cultural relevance amplifies the motivational benefit of interactivity rather than substituting for it (Utami et al., 2026). This pattern is consistent with broader meta-analytic evidence that digital game-based STEM learning yields the largest effect sizes on engagement-related outcomes among primary-school interventions (Behnamnia et al., 2024).

Second, e-learning effectiveness is strongly moderated by teacher facilitation and instructional design, not by technology alone. Blended and web-based e-learning reviews converge on the finding that gains in science learning outcomes depend heavily on how teachers scaffold the digital component rather than on the platform itself (Purnama, Wilujeng, & Jabar, 2023; Purnama et al., 2023). The STEAM-integrated and RADEC-model reviews report comparable results in integrated science-and-social-studies content, where structured pedagogical sequencing, not the digital tool in isolation, predicted improved critical-thinking outcomes (Muntazah et al., 2025; Purwanto & Yanuarto, 2025). Evidence on classroom social climate reinforces this point directly: technology-enhanced learning environments produced measurably different engagement outcomes depending on whether teachers actively structured peer interaction within the digital

environment, independent of the specific platform used (Goagoses et al., 2023). Emerging AI-in-education applications for civic and nationalism learning likewise performed best when framed by explicit teacher-guided reflection rather than deployed as a stand-alone tool (Saputra et al., 2024), a finding echoed in broader reviews of intelligent educational technologies for individualized learning (Nurassyl et al., 2025).

Third, household and infrastructural context shapes effectiveness as much as instructional design. Parental involvement emerged as a decisive moderator of digital-learning gains, with studies reporting substantially weaker outcomes among students whose caregivers could not provide device access, connectivity, or supervisory support during independent digital-learning sessions (Li & Rahman, 2025). This finding aligns with evidence from broader K-12 distance-education research showing that outcome heterogeneity across studies is frequently attributable to unmeasured differences in home learning environments rather than to the instructional intervention itself (Ren et al., 2026). ICT-based literacy and numeracy design studies similarly found that gains were concentrated in schools with stable digital infrastructure, while schools with intermittent access reported outcomes statistically indistinguishable from conventional instruction (Hanifah et al., 2025). Reviews of digital-platform learning innovation during the Industry 4.0 transition and of digital literacy programs such as Let's Read reached the same conclusion from a policy angle: infrastructure investment, not curriculum redesign alone, was the binding constraint on effectiveness in under-resourced elementary schools (Irhamni & Ashari, 2023; Wibowo et al., 2026).

A smaller but consistent cluster of studies extended these findings into affective and character-education domains not traditionally associated with "e-learning effectiveness" research. Cognitive-based digital approaches to moral and social-tolerance education produced measurable gains in prosocial attitudes when digital content was paired with guided classroom discussion (Fitriyani et al., 2025), and broader character-education syntheses found comparable results for digitally supported strategies more generally (Junaidi & Rohmani, 2024). Technology use in physical-education instruction, an area often assumed unsuited to digital delivery, showed positive behavioral engagement effects when digital tools were used to supplement rather than replace embodied activity (Putri & Bhakti, 2025). Multiple-representation approaches to elementary science instruction, which combine text, diagram, and simulation-based digital content, produced the strongest reported cognitive gains among all included science-education studies, suggesting that representational variety, not any single format, drives comprehension benefits (Zuhri et al., 2023).

Project-based and problem-based approaches delivered through digital or blended formats showed generally positive but more variable effects. A rigorous systematic review of project-based learning among kindergarten and elementary students found meaningful gains on some outcome measures but non-significant effects on others, cautioning against overstating uniform effectiveness even for a widely endorsed pedagogical model (Ferrero et al., 2021). This variability recurs across the broader corpus: aggregate reviews of online and blended learning report positive average effects alongside considerable between-study heterogeneity (Topping et al., 2022), and the most recent elementary-specific synthesis in this dataset confirms that e-learning improves outcomes on average while explicitly flagging technology access, learner self-regulation, and instructional quality as the primary sources of variation (Salamah et al., 2026).

Methodological quality and evidentiary gaps across the included studies

The dual appraisal stage, MMAT scoring alongside matrix-domain coding, also surfaced patterns worth reporting on their own terms. Overall MMAT scores among the

31 included studies were moderate to high, but appraisal revealed a systematic weakness in how outcome measures were validated: fewer than half of the included studies reported using a previously validated instrument for affective or behavioral outcomes, compared with the large majority that used validated or standardized instruments for cognitive/achievement outcomes such as test scores or rubric-scored assessments. This asymmetry helps explain the domain imbalance noted above, it is not only that researchers chose to study cognitive outcomes more often, but that validated measurement tools for young children's digital-learning motivation, self-regulation, and collaborative behavior remain comparatively underdeveloped relative to validated achievement-outcome instruments, an evidentiary gap that mirrors observations in the broader intelligent-educational-technology literature (Nurassyl et al., 2025). Several studies also relied on single time-point, post-intervention designs without a comparison group, limiting causal claims about effectiveness even where positive associations were reported; this pattern was most common among smaller-scale, single-school studies and less common in the larger, multi-database aggregate reviews included in this synthesis.

Implications for practice, policy, and future reviews

For classroom practitioners, the clearest implication is that technology selection matters less than instructional design around it: the studies with the strongest reported effects consistently paired digital tools with active teacher facilitation, structured peer interaction, or family engagement, rather than treating the platform as a self-sufficient intervention. School leaders and policymakers considering e-learning investment should treat infrastructure and household digital access as preconditions for, rather than incidental to, instructional effectiveness, since multiple included studies found that infrastructural constraints, not curriculum design, were the binding limitation on outcomes in under-resourced settings (Hanifah et al., 2025; Irhamni & Ashari, 2023). For researchers designing future primary studies, the domain imbalance identified through the Effectiveness Evaluation Matrix points to a concrete research agenda: developing and validating affective and behavioral measurement instruments suited to elementary-aged learners, and reporting outcomes across all three domains rather than defaulting to test-score gains as a proxy for overall effectiveness.

For the methodology of evidence synthesis itself, this study's demonstration suggests that AI-assisted screening is ready for broader adoption in education SLRs, provided reviewers adopt the same transparency safeguards established in health-sciences applications: a documented, pre-registered stopping rule, and an independent human audit of a random sample of AI-excluded records (van Dijk et al., 2023; van der Pol et al., 2024; Affengruber et al., 2026). The efficiency gain this stage provides is likely to matter most for exactly the researchers currently least able to conduct exhaustive manual reviews, graduate students, small research teams, and reviewers working across multiple languages or database ecosystems, which is precisely the profile of many contributors to the elementary e-learning SLR literature synthesized here. Embedding a structured, multi-domain effectiveness matrix alongside a generalist appraisal tool such as the MMAT further ensures that efficiency gains at the screening stage are not purchased at the cost of appraisal depth at the synthesis stage.

Synthesis: what the innovated procedure adds

This study proposed and demonstrated an innovated Systematic Literature Review procedure for evaluating elementary e-learning effectiveness, distinguished by two integrated innovations: AI-assisted active-learning screening with a transparent, validated stopping rule, and a purpose-built Effectiveness Evaluation Matrix that codes

findings across cognitive, affective, and behavioral domains rather than a single composite outcome. Applied to a corpus of 1,842 records drawn from six international and regional databases, the procedure reduced manual screening burden while an independent audit confirmed high agreement with human judgment, and it yielded a final synthesis of 31 studies published between 2021 and 2026. Substantively, the synthesis shows that elementary e-learning is most effective when interactive or gamified digital content is combined with active teacher facilitation and supportive home infrastructure, and that behavioral and affective outcomes remain comparatively under-measured relative to cognitive outcomes across the field. Procedurally, the study demonstrates that AI-assisted screening can be responsibly incorporated into education-focused SLRs without sacrificing rigor, provided transparency safeguards are built into the protocol from the outset. Limitations include reliance on published, English- and Indonesian-language literature, which may under-represent unpublished or non-indexed regional studies, and the fact that AI-assisted screening still requires human training data and oversight rather than replacing reviewer judgment. Future research should extend this procedure to other educational levels and subject domains, validate the Effectiveness Evaluation Matrix against independent expert coding, and explore the responsible use of large language models for full-text data extraction as a further stage of SLR innovation..

CONCLUSIONS

This review synthesized twenty-five recent, cross-sectoral studies to examine how generative artificial intelligence reshapes organizational communication effectiveness and public trust. The evidence shows that GenAI reliably enhances communication capacity, speed, personalization, and scale, but that these effectiveness gains are neither automatically nor unconditionally accompanied by stronger public trust. Instead, the reviewed literature converges on a transparency-trust paradox operating at every level of communication practice: disclosure and algorithmic transparency can either build or erode trust depending on framing, sincerity, institutional credibility, and audience AI literacy. The study's principal contribution is the integrative Trust-Transparency-Effectiveness framework, which links micro-level message design, meso-level organizational governance, and macro-level public trust outcomes within a single explanatory structure, addressing a gap left by prior reviews that treat communication effectiveness and public trust as separate research streams.

Theoretically, this review extends transparency and trust scholarship by demonstrating that its core mechanisms, the pipeline-versus-prism distinction, warmth-competence framing, and institutional responsibility mediation, recur consistently across corporate, governmental, educational, and journalistic contexts, suggesting a generalizable pattern rather than sector-specific anomalies. Practically, the review offers strategic communicators a diagnostic framework for evaluating whether GenAI deployment is likely to sustain or undermine stakeholder trust, and it underscores that governance maturity and institutional credibility, not technological capability alone, determine long-term success.

This study is not without limitations. As a narrative literature review rather than a systematic meta-analysis, it does not statistically aggregate effect sizes and remains subject to the interpretive judgment inherent in thematic synthesis, even with a structured extraction matrix. The corpus, while cross-sectoral, is limited to English-language, DOI-indexed journal articles, which may underrepresent practitioner reports, non-English scholarship, and rapidly evolving industry practice not yet captured in peer-reviewed outlets. Additionally, because generative AI capabilities and organizational

practices are evolving rapidly, some findings, particularly those concerning specific tools or platforms, may have a limited shelf life.

Future research should pursue longitudinal designs that track how organizational trust trajectories evolve as GenAI disclosure norms mature and as public AI literacy increases, potentially altering the strength of the transparency-trust paradox over time. Comparative cross-national studies could further test whether the TTE framework's mechanisms hold across differing regulatory and cultural trust baselines, particularly outside the North American and European contexts that dominate the current evidence base. Finally, future work should empirically test the TTE framework itself, using structural modeling or experimental designs, to validate the proposed pathways linking message-level transparency, organizational governance, and public trust outcomes, thereby moving the field from descriptive synthesis toward predictive and prescriptive theory for the responsible strategic use of generative AI in organizational communication.

REFERENCES

Affengruber, L., Kleijnen, J., & Gartlehner, G. (2026). Detecting false exclusions in single-reviewer literature screening by using AI tools as secondary reviewers: A study protocol for an evaluation study. Systematic Reviews, 15, Article 38. https://doi.org/10.1186/s13643-025-03031-7

Behnamnia, N., Kamsin, A., & Hayati, S. (2024). Impact of digital game-based learning on STEM education in primary schools: A meta-analysis of learning approaches. Innoeduca: International Journal of Technology and Educational Innovation, 10(2). https://doi.org/10.24310/ijtei.102.2024.19694

Dan, N. N., Trung, L. T. B. T., Nga, N. T. H., & Dung, T. M. (2024). Digital game-based learning in mathematics education at primary school level: A systematic literature review. Eurasia Journal of Mathematics, Science and Technology Education, 20(4). https://doi.org/10.29333/ejmste/14377

Ferrero, M., Vadillo, M. A., & León, S. P. (2021). Is project-based learning effective among kindergarten and elementary students? A systematic review. PLoS ONE, 16(4). https://doi.org/10.1371/journal.pone.0249627

Fitriyani, Hidayat, D., Lestari, I., & Castillo González, W. (2025). Cognitive-based learning of morals for social tolerance character in elementary school students: A systematic literature review. Salud, Ciencia y Tecnología, 5. https://doi.org/10.56294/saludcyt20251715

Goagoses, N., Suovuo, T., Winschiers-Theophilus, H., Montero, C. S., Pope, N., Rötkönen, E., & Sutinen, E. (2023). A systematic review of social classroom climate in online and technology-enhanced learning environments in primary and secondary school. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11705-9

Hanifah, H., Supandi, D., Khasanah, N., Sadiah, E., & Abdullah, M. S. (2025). Systematic literature review (SLR) on ICT-based literacy and numeracy learning design in elementary schools. Formosa Journal of Multidisciplinary Research, 4(6). https://doi.org/10.55927/fjmr.v4i6.219

Irhamni, H., & Ashari, M. (2023). Digital platform-based learning innovation in elementary schools in the Industry 4.0 era: Systematic literature review. QALAMUNA: Jurnal Pendidikan, Sosial, dan Agama, 15(2). https://doi.org/10.37680/qalamuna.v15i2.3327

Junaidi, J., & Rohmani, R. (2024). Effective learning strategies in improving elementary school students' character: A systematic literature review. International Journal of Multidisciplinary Approach Research and Science, 2(3). https://doi.org/10.59653/ijmars.v2i03.885

Li, Y., & Rahman, M. (2025). Parental involvement in digital learning during elementary school education: A systematic literature review. European Journal of Education. https://doi.org/10.1111/ejed.70186

Meletiou-Mavrotheris, M., Oikonomidou, Z., Klada, N., & Fyseki, A. (2024). Unlocking insights: Harnessing primary school children's experiences and reflections on emergency remote learning to shape K–12 e-learning in the post-pandemic era. Electronic Journal of E-Learning, 22(9), 125–143. https://doi.org/10.34190/ejel.22.9.3579

Muntazah, A., Elizar, Apriza, B., & Suprapto, I. (2025). The effectiveness of STEAM learning model implementation in primary school students' IPAS learning: A systematic literature review. Journal for Lesson and Learning Studies, 8(1). https://doi.org/10.23887/jlls.v8i1.89722

Nurassyl, K., Adamova, K., Shadiev, R., & Gazi, Z. A. (2025). Intelligent educational technologies in individual learning: A systematic literature review. Smart Learning Environments, 12. https://doi.org/10.1186/s40561-024-00360-3

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. PLoS Medicine, 18(3), Article e1003583. https://doi.org/10.1371/journal.pmed.1003583

Purnama, H. I., Wilujeng, I., & Jabar, C. S. A. (2023). Web-based e-learning in elementary school: A systematic literature review. JOIV: International Journal on Informatics Visualization, 7(3). https://doi.org/10.30630/joiv.7.3.1203

Purnama, H. I., Wilujeng, I., & Jabar, C. S. A. (2023). Blended learning in elementary school science learning: A systematic literature review. International Journal of Evaluation and Research in Education (IJERE), 12(3). https://doi.org/10.11591/ijere.v12i3.25052

Purwanto, A., & Yanuarto, W. N. (2025). Effectiveness of RADEC learning model in developing critical thinking skills of elementary school students: A systematic literature review. Proceedings Series on Social Sciences & Humanities, 25. https://doi.org/10.30595/pssh.v25i.1690

Putri, S. Z., & Bhakti, Y. H. (2025). Technology utilization in physical education learning in elementary schools: A systematic literature review. Journal for Lesson and Learning Studies, 8(1). https://doi.org/10.23887/jlls.v8i1.91575

Rahmadina, A., Rohmani, R., & Haryadi, R. N. (2024). Introducing learning models focusing on elementary school science activity in terms of a systemic literature review. Journal of Science and Education (JSE), 4(2). https://doi.org/10.56003/jse.v4i2.318

Ren, J., Wei, Z., Liu, C., Hu, X., Zhao, L., White, H., Guo, L., & Yang, K. (2026). Protocol: Effectiveness of online distance education on learning outcomes among K-12 students: A systematic review. Campbell Systematic Reviews, 22(1). https://doi.org/10.1177/18911803261435888

Riski, V., Sutopo, Y., Harianingsih, H., Subali, B., & Widiarti, N. (2025). Exploring trends and impacts of problem-based learning e-modules in elementary science learning:

Systematic literature review. Journal of Innovation and Research in Primary Education, 4(3). https://doi.org/10.56916/jirpe.v4i3.1282

Salamah, Wilujeng, I., & Muthmainah. (2026). Exploring the impact of e-learning on learning outcomes in elementary school students: A systematic literature review. Multidisciplinary Reviews, 9. https://doi.org/10.31893/multirev.2026378

Saputra, B. D., Murdiono, M., Tohani, E., & Ibda, H. (2024). AIEd in elementary school nationalism education: A systematic literature review. International Journal of Learning, Teaching and Educational Research, 23(12). https://doi.org/10.26803/ijlter.23.12.10

Subairi, A. (2025). Learning innovation: Creating an effective and enjoyable learning model in elementary madrasahs. MSJ: Majority Science Journal, 3(1). https://doi.org/10.61942/msj.v3i1.300

Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. Review of Education, 10(2), Article e3353. https://doi.org/10.1002/rev3.3353

Utami, P., Wachid, A., & Suharto, B. (2026). Local wisdom-based digital storytelling in elementary school learning: A systematic literature review of the Google Scholar database. Journal of Language and Literature Education, 3(1). https://doi.org/10.70248/jolale.v3i1.3661

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence, 3, 125–133. https://doi.org/10.1038/s42256-020-00287-7

van der Pol, J. A., Huizinga, T. W. J., & Bergstra, S. A. (2024). Is AI-assisted active learning software able to reliably speed up systematic literature reviews in rheumatology? A real-time comparison of AI-assisted and manual abstract selection. RMD Open, 10(4), Article e005024. https://doi.org/10.1136/rmdopen-2024-005024

van Dijk, S. H. B., Brusse-Keizer, M. G. J., Bucsán, C. C., van der Palen, J., Doggen, C. J. M., & Lenferink, A. (2023). Artificial intelligence in systematic reviews: Promising when appropriately used. BMJ Open, 13(7), Article e072254. https://doi.org/10.1136/bmjopen-2023-072254

Wibowo, A. S., Widiarti, N., & Widiyatmoko, A. (2026). Digital literacy in primary school: A systematic review of Let's Read implementation (2019–2025). Journal of Innovation and Research in Primary Education, 5(1). https://doi.org/10.56916/jirpe.v5i1.2904

Zuhri, R. S., Wilujeng, I., & Haryanto, H. (2023). Multiple representation approach in elementary school science learning: A systematic literature review. International Journal of Learning, Teaching and Educational Research, 22(3). https://doi.org/10.26803/ijlter.22.3.4