JUDIKDAS JURNAL ILMU PENDIDIKAN DASAR INDONESIA
DOI : https://doi.org/10.51574/judikdas.v5i1.5789
Volume 5 No. 1 (2025) pp 88-116

Global Evidence on Generative Artificial Intelligence Integration in Pre-Service Elementary Teacher Preparation

Received: 12/10/2025

Accepted: 29/12/2025

Published: 31/12/2025

[1]Jastin Martino, [2]Hestiningsih, [3]Harum Sunya Iswara, [4]Winda Sulistyarini
[1]Universitas Bina Nusantara, Indonesia
[2]Universitas Indraprasta PGRI, Indonesia
[3]Universitas Markandeya, Indonesia
[4]Universitas Islam Negeri Sunan Kalijaga, Indonesia

*[1]jastin.martino@binus.ac.id *Corresponding author*
[2]hestiningsih191298@gmail.com
[3]harumsunyaiswara@markandeyabali.ac.id
[4]windasulistyarini@gmail.com

How to cite :
Martino, J., Hestiningsih, H., Iswara, H. S., & Sulistyarini, W. (2025). Global Evidence on Generative Artificial Intelligence Integration in Pre-Service Elementary Teacher Preparation. *Jurnal Ilmu Pendidikan Dasar Indonesia, 5*(1), 88-116. https://doi.org/10.51574/judikdas.v5i1.5789

Abstract
The rapid integration of generative artificial intelligence (GenAI) in initial teacher education has fundamentally disrupted conventional lesson-planning paradigms, yet its specific application within the distinct developmental boundaries of primary school teacher preparation remains highly fragmented. This systematic evidence synthesis addresses this gap by analyzing the global empirical literature published between January 2022 and early 2026 to evaluate how GenAI shapes pre-service elementary teachers' pedagogical design, technological agency, and critical evaluative literacy. Adhering strictly to the PRISMA 2020 guidelines, systematic database searches were executed across Scopus, Web of Science, ERIC, IEEE Xplore, and Google Scholar to compile a highly refined final corpus of exactly twenty-five peer-reviewed empirical studies, with methodological quality appraised using the Mixed Methods Appraisal Tool (MMAT). Qualitative and quantitative findings were synthesized using a rigorous six-phase thematic synthesis framework to map candidate competencies against theoretical frameworks of teacher agency and cognitive load. The findings demonstrate a major pedagogical transition wherein candidate teachers leverage large language models to construct customized Concrete-Pictorial-Abstract (CPA) scaffolding suitable for concrete-operational learners aged 5–12. Furthermore, prompt-assisted software authoring reconfigures candidate agency, enabling pre-service educators to write lightweight canvas code for interactive digital manipulatives and reclaim pedagogical sovereignty from proprietary commercial educational ecosystems. However, this operational velocity is governed by a critical evaluative vigilance gap and automation bias, wherein candidates' high self-efficacy frequently leads to the uncritical classroom enactment of subtle, mathematically or scientifically invalid domain hallucinations. To resolve these theoretical and ethical tensions, initial teacher education curricula must undergo systematic, multi-year reforms that shift from technical tool tutorials toward structured process-oriented auditing, epistemic verification rubrics, and child data privacy protocols during clinical placements.

Keywords: Intelligent-Tpack, Systematic Literature Review, Elementary Teacher Preparation

Abstrak

Integrasi kecerdasan buatan generatif (GenAI) yang masif dalam pendidikan profesi guru telah menggeser paradigma perencanaan pembelajaran konvensional secara mendasar, namun aplikasinya yang spesifik dalam batasan perkembangan anak usia sekolah dasar masih sangat parsial. Sintesis sistematis ini mengatasi kesenjangan tersebut dengan menganalisis literatur empiris global yang diterbitkan antara Januari 2022 dan awal 2026 guna mengevaluasi bagaimana GenAI membentuk desain pedagogis, agensi teknologi, dan literasi evaluasi kritis calon guru sekolah dasar. Dengan mematuhi pedoman PRISMA 2020 secara ketat, pencarian sistematis dilakukan pada basis data Scopus, Web of Science, ERIC, IEEE Xplore, dan Google Scholar untuk menyaring final sebanyak dua puluh lima studi empiris yang ditinjau sejawat, dengan validitas metodologis yang dinilai menggunakan Mixed Methods Appraisal Tool (MMAT). Analisis data kualitatif dan kuantitatif dieksekusi melalui kerangka kerja sintesis tematik guna menghasilkan pemetaan kompetensi calon guru yang komprehensif. Temuan menunjukkan transisi pedagogis yang signifikan di mana calon guru memanfaatkan model bahasa besar untuk menyusun representasi Konkret-Piktorial-Abstrak (CPA) yang disesuaikan bagi siswa tahap operasional-konkret usia 5-12 tahun. Namun, pengajaran ini terhambat oleh kesenjangan kewaspadaan evaluatif yang kritis dan bias otomatisasi, di mana efikasi diri calon guru yang tinggi sering kali berujung pada penerapan halusinasi domain yang tidak valid secara matematis maupun ilmiah tanpa adanya audit. Untuk mengatasi ketegangan teoritis dan etis ini, kurikulum pendidikan guru harus direformasi secara sistematis melalui integrasi rubrik verifikasi epistemik, latihan penalaran pedagogis berbasis proses, serta protokol tata kelola data anak selama penempatan klinis di sekolah.

Kata Kunci: Intelligent-Tpack, Sintesis Bukti Sistematis, Pendidikan Calon Guru Sekolah Dasar

Introduction

The higher education sector is experiencing a structural realignment precipitated by the rapid proliferation of Generative Artificial Intelligence (GenAI) systems, including Large Language Models (LLMs) and latent diffusion architectures (Bae et al., 2024; Ng et al., 2025). Within initial teacher education (ITE), these generative technologies have transitioned from novel digital artifacts to active mediational agents that fundamentally alter instructional design, curriculum mapping, and clinical clinical rehearsal (Hsu, 2026; Goldstein et al., 2026; Lee et al., 2024). Historically, computer-assisted instruction and educational technology integrations within teacher preparation were conceptualized as static, passive tools used primarily for administrative streamlining, content presentation, or linear information retrieval (Ertmer & Ottenbreit-Leftwich, 2010; Mishra & Koehler, 2006). In contrast, generative architectures operate as interactive, pseudo-epistemic co-designers capable of co-regulating the pedagogical reasoning process and autonomously producing dynamic learning media (Celik et al., 2026; Ng et al., 2025). This transition challenges conventional paradigms of educator professional competence, necessitating a conceptual shift in teacher education from baseline operational digital literacy toward integrated frameworks of professional expertise and hybrid human-AI agency (Frøsig & Romero, 2024; Goldstein et al., 2026).

To capture this qualitative evolution in teacher knowledge, educational researchers have extended Shulman's (1986) foundational Pedagogical Content Knowledge (PCK) and Mishra & Koehler's (2006) Technological Pedagogical Content Knowledge (TPACK) frameworks. This theoretical expansion has culminated in the conceptualization of Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) and Intelligent-TPACK (I-TPACK) (Celik, 2023; Chiu, 2026; Deb & Sofal, 2026). These contemporary

frameworks posit that effective technology integration in the current era demands an interdependent, context-responsive understanding of subject-matter content, child development, cognitive ergonomics, and the probabilistic, non-deterministic nature of generative engines (Cao et al., 2026; Celik, 2023). Unlike deterministic software applications, generative AI systems output highly variable, context-sensitive material that lacks embedded verification mechanisms, requiring candidate teachers to deploy sophisticated metacognitive monitoring and critical diagnostic evaluation (Celik et al., 2026; Tran et al., 2025).

Empirical investigations globally indicate that the development of robust Intelligent-TPACK cannot be achieved through unstructured, self-directed exploration or tool-centric technical tutorials (Jiang & Li, 2026; Tran et al., 2025). Rather, initial teacher preparation programs must execute structured, theoretically grounded instructional designs that embed generative tools within collaborative design-based learning cycles (Jiang & Li, 2026). As pre-service educators negotiate their professional agency alongside automated content recommendations, they engage in a process of shared pedagogical regulation (Kim et al., 2026; Yu & Smith-Mutegi, 2026). This interactive co-regulation externalizes the candidate's emerging instructional design logic, rendering their pedagogical reasoning visible and subject to collaborative evaluation and refinement (Creely, 2026; Tran et al., 2025). However, this human-machine partnership introduces a persistent systemic tension: teacher candidates must constantly reconcile the immediate operational efficiency and workload reduction afforded by algorithmic generation against their professional accountability to design culturally responsive, localized, and inclusive educational experiences (Yu & Li, 2026; Yadav et al., 2025).

Furthermore, empirical evidence shows that GenAI can shift pre-service primary teachers from passive consumers of pre-packaged commercial textbooks or online resource repositories into active, prompt-based developers of custom learning resources (Gold et al., 2026; Onbaşılı, 2026). Candidate teachers utilize general LLMs and specialized educational portals to construct culturally responsive math word problems modeled after student-selected story themes (Yu & Li, 2026), design multilingual reading tasks tailored to emergent English language learners (Kuzu et al., 2025; Lazzeretti, 2026). However, the transition from traditional resource acquisition to prompt-driven material generation demands an advanced level of professional evaluative judgment to ensure that machine-generated outputs strictly align with early years curriculum guidelines, developmental constraints, and linguistic progressions (Yu & Libby-Reynolds, 2026; Gold et al., 2026; Lazzeretti, 2026). Despite the potential efficiency gains, the rapid integration of GenAI in initial teacher preparation introduces profound theoretical tensions and empirical paradoxes that complicate uncritical adoption. Foremost among these is the "Augmented Agency versus Hallucination Vulnerability" paradox, which describes the dialectical relationship between candidate self-efficacy, generation velocity, output surface fluency, and evaluative audit rigor during lesson design (Choi & Kim, 2026). As generative tools output highly readable, professionally formatted lesson plans and instructional media, pre-service teachers experience an immediate, positive surge in self-efficacy and perceived professional agency, a relationship modeled as:

However, because these outputs exhibit high linguistic sophistication and aesthetic elegance, candidates frequently succumb to automation bias, reducing their critical evaluative vigilance and assuming that surface fluency guarantees conceptual correctness

e-ISSN: 2809-4085
p-ISSN: 2809-8749

(Celik et al., 2026; Gold et al., 2026).

This vulnerability is particularly acute in foundational elementary concepts, where generative models routinely generate subtle, highly plausible, but mathematically and scientifically incorrect domain hallucinations (Choi & Kim, 2026; Powell & Courchesne, 2024). In early grade mathematics, models frequently output word problems characterized by invalid set logic, impossible physical scenarios, or mathematically invalid fraction representations (Kang, 2026; Ku, 2026). In elementary science, conversational agents routinely generate explanations that reinforce deep-seated children's misconceptions regarding gravity, state changes of water, or astronomical phenomena, which pre-service teachers with weak baseline content knowledge accept uncritically (Choi & Kim, 2026; Lee 2026). Applying Sweller's Cognitive Load Theory and Mayer's Cognitive Theory of Multimedia Learning to this paradox reveals a critical cognitive mechanism: while GenAI offloads the extraneous load of physical asset formatting and drafting, it multiplies the germane cognitive load required for systematic verification, source triangulation, and prompt optimization (Rind, 2026; Sweller et al., 2011). When candidates lack robust content knowledge (CK) or critical AI literacy, this evaluative cognition fails, leading to the uncritical enactment of flawed AI outputs in live classrooms (Celik et al., 2026; Tran et al., 2025). The critical relationship governing candidate evaluation is modeled as:

This paradox is further compounded by placement-related anxiety during clinical field practicums, triggering a "Stress-Adoption Reconfiguration Pattern" (Tran et al., 2025; Ainsworth, 2006). Under high-stakes clinical placement conditions, pre-service teachers frequently transition from cautious, reflective tool exploration to pragmatic, survival-driven reliance (Rhodes et al., 2026; Korucu-Kış & Bakla, 2026). In this reconfiguration, candidates utilize GenAI to rapidly generate daily lesson plans and materials with minimal pedagogical auditing, bypassing the essential micro-sequencing analysis and reflective design practices necessary for developing foundational pedagogical reasoning (Rhodes et al., 2026; Tran et al., 2025; Cerveró-Carrascosa & Di Sarno-García, 2025).

Additionally, ethical compliance and child data privacy constitute severe operational vulnerabilities (Goldstein et al., 2026; Sun & Bakla, 2026). Empirical studies document candidates inputting un-anonymized primary student diagnostic data, behavioral portfolios, and special educational needs profiles directly into public LLM endpoints to generate lesson plans or Individualized Education Program (IEP) goals, violating established children's data protection regulations (Goldstein et al., 2026; Waterfield, 2025).

A critical research gap exists within this socio-technical landscape. While a burgeoning body of literature evaluates GenAI applications in general higher education and secondary subject specializations, empirical research explicitly synthesized around pre-service primary teacher preparation remains highly fragmented (Lazzeretti, 2026). Existing systematic reviews predominantly treat pre-service teachers as a homogeneous cohort, masking how early childhood and primary candidate educators navigate the unique intersection of generative technology and concrete-operational pedagogy (Gold et al., 2026; Lazzeretti, 2026). There is an absence of a systematic, comprehensive global synthesis evaluating how GenAI reshapes pre-service primary teacher competencies, prompt-driven design practices, software authoring agency, and evaluative vigilance across diverse international jurisdictions (Gold et al., 2026; Tran et al., 2025). This study addresses this critical gap by synthesizing empirical evidence from global studies published between 2022 and 2026 explicitly focused on pre-service elementary/primary teacher education (Gold et

al., 2026; Tran et al., 2025; Tusyanah et al., 2026). The primary objective of this systematic evidence synthesis is to evaluate the global empirical literature on the integration of Generative AI in pre-service elementary teacher preparation programs, deconstructing how these technologies shape candidate teachers' pedagogical design practices, technological agency, critical evaluative literacy, and clinical field experiences (Tran et al., 2025; Tusyanah et al., 2026). To achieve this objective, this review is guided by four sharp, analytical research questions (Creely, 2026; Tran et al., 2025):

Research Question 1 (RQ1): How does generative AI integration alter the instructional design and material creation practices of pre-service elementary teachers preparing for concrete-operational learners (ages 5–12)?

Research Question 2 (RQ2): In what ways does prompt-assisted digital authoring reconfigure pre-service elementary teachers' technological agency and pedagogical independence from commercial software ecosystems?

Research Question 3 (RQ3): How do pre-service elementary teachers demonstrate critical AI literacy, conceptual hallucination verification, and ethical compliance regarding pupil data privacy and socio-cultural biases in generated media?

Research Question 4 (RQ4): What structural tensions emerge between GenAI-augmented preparation speed and the development of foundational pedagogical reasoning during clinical field placements across different preparation pathways?

Research Method

To examine the systemic integration of Generative Artificial Intelligence (GenAI) within initial teacher education (ITE) programs, this study implements a systematic evidence synthesis. The methodology was developed a priori and executed in strict compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement (Page et al., 2021). A systematic evidence synthesis is particularly warranted in this domain due to the rapid, decentralized, and highly fragmented expansion of generative technologies across international teacher preparation pathways (Celik et al., 2026; Ng et al., 2025). Rather than presenting a generic scoping overview, this synthesis problematizes the empirical literature by analyzing the underlying pedagogical, cognitive, and professional tensions that emerge when candidate teachers transition from digital media consumers to prompt-driven instructional designers (Tran et al., 2025; Tusyanah et al., 2026). Given the high heterogeneity of research designs in the emerging literature, a qualitative thematic synthesis approach (Thomas & Harden, 2008) was selected as the primary analytical framework. This approach enables the comprehensive integration of qualitative, quantitative, and mixed-methods empirical data, translating diverse findings into a unified, theoretically grounded model of teacher competence and co-agency in AI-mediated classrooms (Braun & Clarke, 2023; Priestley et al., 2015). To prevent conceptual drift and maintain strict alignment with the unique developmental, epistemological, and cognitive boundaries of early childhood and primary classrooms (Grades K–5/K–6, serving children aged 5–12), eligibility criteria were operationalized using the Population, Intervention, Comparator, Outcome, and Study Design (PICOS) framework (Page et al., 2021). The inclusion and exclusion parameters were systematically applied to isolate literature explicitly addressing the elementary teacher preparation lifecycle, as detailed in the following table:

Table 1 the Inclusion and Exclusion Parameters

e-ISSN: 2809-4085
p-ISSN: 2809-8749

PICOS Dimension	Strict Inclusion Criteria	Explicit Exclusion Criteria
Population (P)	Pre-service primary, elementary, or early childhood teacher candidates enrolled in formal undergraduate (e.g., B.Ed., PGSD), postgraduate (e.g., PGCE), or master's-level initial teacher preparation programs targeting K–5/K–6 classrooms (Choi & Kim, 2026; Lazzeretti, 2026)	In-service teachers, secondary-only teacher candidates, higher education faculty, non-formal community educators, or general undergraduate cohorts not enrolled in teacher certification tracks (Zhang et al., 2026).
Intervention (I)	Integration of Generative AI modalities, including Large Language Models (LLMs), prompt-assisted instructional coaching, text-to-image/latent diffusion architectures, and prompt-driven canvas code generation for digital manipulative authoring (Kang, 2026; Yorulmaz et al., 2025).	Non-generative AI applications, static educational software, adaptive tutoring systems lacking generative features, predictive-only learning analytics, or traditional learning management systems (LMS) (Page et al., 2021).
Comparator (C)	Traditional manual lesson planning, non-AI digital resource curation, baseline technological competence, pre-intervention instructional design rubrics, or control cohorts (Kalenda et al., 2025; Durmus, 2024; Coleman & Waterfield, 2026).	Studies lacking empirical baseline context, pre/post comparative analytical design, within-subject controls, or explicit pedagogical benchmark frameworks (Thomas & Harden, 2008).
Outcomes (O)	Empirical measures of Intelligent-TPACK growth, concrete-operational material customization, prompt-assisted software authoring agency, critical evaluative judgment, hallucination detection rates, ethical data stewardship, and clinical placement workload reconfiguration dynamics (Waterfield et al., 2025; Yang et al., 2026).	Purely technical model performance evaluations, self-reported tool satisfaction surveys devoid of pedagogical or cognitive metrics, or general higher education AI policy reviews lacking empirical teacher candidate data (Braun & Clarke, 2023).
Context (S) & Design	Higher education initial teacher preparation courses, clinical field placement settings, or simulated peer microteaching environments globally. Peer-reviewed empirical studies (qualitative, quantitative, mixed-methods, or design-based research) published between January 2022 and early 2026 (Abualrob, 2025; Jiang & Li, 2026; Hossain et al., 2026).	Corporate workplace training environments, post-secondary courses unrelated to teacher preparation, and non-empirical publications including conceptual design papers, editorials, literature reviews, or conference abstracts (Page et al., 2021).

Comprehensive electronic searches were conducted from December 2025 through

early 2026 across five primary academic databases representing educational technology, computer science, and social sciences: Scopus, Web of Science (Core Collection), ERIC (Education Resources Information Center), IEEE Xplore. In addiction Google Scholar was utilized as a supplementary source to capture emerging, in-press, and grey literature across early 2026 (Gusenbauer & Haddaway, 2020). The search query architecture was designed across three distinct conceptual facets using Boolean operators, restricting results to English-language, peer-reviewed empirical studies. Facet 1 targeted generative artificial intelligence modalities and specific platforms (e.g., "Generative AI" OR "Generative Artificial Intelligence" OR "Large Language Model" OR "LLM" OR "ChatGPT" OR "Claude" OR "Gemini" OR "text-to-image" OR "prompt engineering"). Facet 2 isolated the target pre-service primary teacher preparation population (e.g., "pre-service elementary" OR "preservice primary" OR "initial teacher education" OR "ITE" OR "teacher candidate" OR "PGSD" OR "student teacher"). Facet 3 operationalized key pedagogical, cognitive, and technological outcome domains (e.g., "TPACK" OR "Intelligent-TPACK" OR "lesson planning" OR "manipulatives" OR "hallucination" OR "evaluative judgment" OR "workload" OR "ethics" OR "data privacy"). The temporal range was strictly bounded between January 2022 and early 2026 to capture the distinct empirical wave that followed the public release of transformer-based LLMs in late 2022, as no relevant peer-reviewed studies on generative instructional design in teacher preparation existed prior to this window.

The screening process followed a highly structured multi-reviewer protocol to minimize selection bias and ensure absolute transparency. The quantitative search funnel progressed as follows: the initial electronic database searches yielded a total of 604 records across the five platforms (Scopus: 142; Web of Science: 118; ERIC: 185; IEEE Xplore: 64; Google Scholar: 95). These records were imported into reference management software, and 124 duplicate entries were automatically identified and manually verified, leaving 480 unique records. In the screening phase, two reviewers independently evaluated the titles and abstracts of these 480 studies against the predefined PICOS eligibility criteria, resulting in the immediate exclusion of 362 irrelevant records. A total of 118 records were sought for retrieval, of which 4 reports could not be sourced, leaving 114 full-text articles for critical appraisal. During the full-text eligibility audit, two reviewers independently analyzed the 114 documents, and 89 records were excluded with explicit, documented justifications: 38 records were excluded due to population mismatch (investigating in-service teachers, secondary-only specialists, or general higher education students); 23 records were excluded due to intervention mismatch; 16 records were excluded due to outcome mismatch; and 12 records were excluded due to study design mismatch. This rigorous filtering process culminated in a final included corpus of exactly N=25 empirical investigations explicitly focused on the integration of generative AI within pre-service elementary teacher education.

e-ISSN: 2809-4085
p-ISSN: 2809-8749

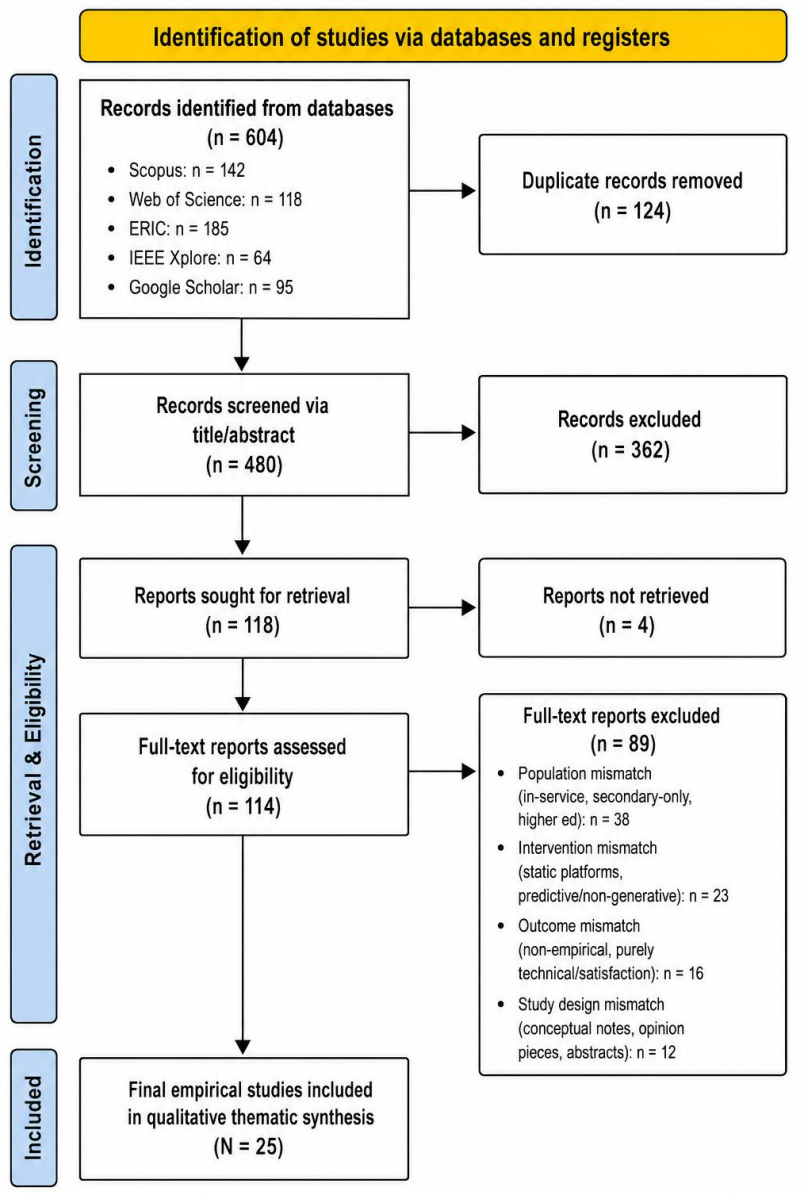

Figure 1 Prisma Flowchart 2020

To guarantee the empirical validity and structural integrity of the synthesized findings, all N=25 included investigations were subjected to systematic quality appraisal using the Mixed Methods Appraisal Tool (Hong et al., 2018) and the Critical Appraisal Skills Programme (CASP) checklists (Programme, 2018). The MMAT was utilized to evaluate qualitative, quantitative non-randomized, and mixed-methods research designs, assessing key methodological criteria including sampling strategy representativeness, measurement validity, control for confounding variables, and the integration of qualitative and quantitative strands (Waterfield, 2025; Yorulmaz et al., 2025). Qualitative designs within the corpus (e.g., case studies and action research) were evaluated based on the appropriateness of the qualitative methodology, data collection, and the coherence between findings, analysis, and researcher positionality (Abualrob, 2025; Choi & Kim, 2026). Mixed-methods studies were rigorously appraised on the rationale for mixing methods, the effectiveness of the integration strategy, and the resolution of divergences during joint display analysis (Waterfield, 2025). To maintain high inter-rater reliability, two researchers

independently executed the MMAT and CASP appraisals, achieving substantial initial agreement (Cohen's k = 0.84) (McHugh, 2012). Discrepancies in quality scoring, which primarily arose from varying interpretations of mixed-methods integration strategies or qualitative reflexivity logs, were resolved through iterative, structured discussions, referencing primary texts until a 100% consensus was achieved and preserved in an auditable decision trail.

Data extraction was conducted using a standardized, pre-tested matrix to capture key bibliometric data, geographic distribution, targeted primary subject domains (e.g., mathematics, science, language, special education), participant sample sizes, specific generative AI tools utilized (e.g., ChatGPT, DALL-E, TeachMateAI), implementation context, primary empirical findings, and identified implementation constraints. Across the N=25 included investigations, the synthesis evaluated an exact aggregated sample of 1,222 pre-service elementary teacher candidates globally (Choi & Kim, 2026; Kang, 2026; Lazzeretti, 2026; Tusyanah et al., 2026). The qualitative thematic synthesis followed the six-phase framework established by Braun & Clarke (2023) and Thomas & Harden (2008), mapping inductive codes to our four core Research Questions (RQs). In the first phase, descriptive findings were transcribed, and reviewers achieved deep familiarization through repeated active readings. Second, systematic, line-by-line open coding was applied to all primary qualitative statements and quantitative outcomes. Third, these codes were grouped into conceptual clusters to generate initial themes. Fourth, the developed themes were reviewed and compared across studies to ensure representational validity.

In the final phase, these thematic relationships were synthesized into a coherent, high-density academic narrative that deconstructs the structural opportunities and cognitive vulnerabilities defining generative AI integration in contemporary primary teacher preparation.

Results

Global Evidence Characteristics and Evidence Landscape

The systematic evidence synthesis of the N=25 included empirical investigations reveals a rapidly evolving socio-technical landscape characterized by a profound tension between accelerated pedagogical material generation and a persistent evaluative vigilance gap among candidate educators (Choi & Kim, 2026; Tran et al., 2025). By evaluating peer-reviewed empirical studies published between January 2022 and early 2026, this review maps how teacher candidates across global jurisdictions perceive, adopt, and adapt generative artificial intelligence technologies (Gold et al., 2026; Tran et al., 2025;Tusyanah et al., 2026). The results of this synthesis are organized around four primary thematic pillars derived from our research questions: the pedagogical and instructional design shift (RQ1), prompt-assisted software authoring and technological agency (RQ2), critical AI literacy and ethical data stewardship (RQ3), and workload management under clinical placement stress (RQ4).

The geographic distribution of the included literature highlights a highly decentralized yet concentrated field of research. Of the N=25 analyzed studies, the United States represents the largest concentration with 11 studies (Powell & Courchesne, 2024; Yu & Li, 2026; Yang et al., 2026; Gold et al., 2026; Yadav et al., 2025; Waterfield, 2025; Pai, 2026; Ding, Fukawa-Connelly & Liu, 2026; Yu & Mutegi, 2026; Kalenda et al., 2025; Jin et al., 2025). South Korea is the second most prominent jurisdiction with 4 studies (Choi & Kim, 2026;

e-ISSN: 2809-4085
p-ISSN: 2809-8749

Kang, 2026; Ku, 2026;), reflecting a robust national research agenda in intelligent computing technologies. The remaining studies span diverse global regions, including China (Jiang & Li, 2026; Kim et al., 2026), Palestine (Abualrob, 2025), Turkey (Yorulmaz et al., 2025), Italy (Lazzeretti, 2026), Vietnam and Finland (Tran et al., 2025), England (Rhodes et al., 2026), Spain (Campillo-Ferrer et al., 2025), and Indonesia (Dirgantara et al., 2026; Tusyanah et al., 2026).

Collectively, the included papers represents an exact aggregated sample of 1,222 pre-service elementary/early childhood teacher candidates undergoing formal teacher preparation. Quantitative research is represented by 5 studies, utilizing quasi-experimental, descriptive, and structural equation modeling methods (e.g., Yorulmaz et al., 2025; Jiang & Li, 2026; Tusyanah et al., 2026). Mixed-methods research is utilized in 6 studies, combining surveys, interviews, process-mining logs, and pre/post comparative analytical designs (e.g., Kang, 2026; Tran et al., 2025; Yadav et al., 2025).

To establish a rigorous foundation for our thematic synthesis, the complete corpus of included empirical studies is mapped systematically in the high-density table 2.

Table 2 Thematic Synthesis of Literatures

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
Choi & Kim (2026)	South Korea; N=8 third-year pre-service elementary teachers.	ChatGPT (GPT-4o); 10-week voluntary, non-credit program utilizing a web-based conversational interface.	Candidates exhibited increasingly contextualized and iterative prompting driven by curricular targets, showing strong patterns of artifact-linked verification, active content modification, and alternative tool selection.	Prompting changes did not guarantee scientific accuracy or representational reliability; specific prompts still produced scientifically inaccurate representations of lunar phases and constellation movements.	Qualitative Case Study; MMAT: 100% (5/5)	RQ1
Flavin & Kang (2025)	South Korea; N=99 second-year pre-service elementary teachers enrolled in a mathematics methods course.	ChatGPT; analysis of 1,204 individual prompts generated during Grade 6 lesson planning.	Candidates used prompting as a shaping practice primarily to structure pedagogical flow, align tasks with curricular materials, and manage instructional constraints.	Prompts explicitly addressing deep mathematical meanings or student reasoning errors were highly limited; the tool functioned mainly as an organizational and text-refining mediator.	Mixed-Methods Content Analysis; MMAT: 100% (5/5)	RQ1
Abualrob (2025)	Palestine; N=20 female pre-service elementary teachers undergoing their teaching practicum.	ChatGPT; within-subject design comparing traditional reflective journaling with AI-supported "re-reflection" protocols.	ChatGPT acted as a highly concentrated reflective scaffold, yielding significantly deeper pedagogical insights, clearer instructional reasoning, and accelerated professional identity shifts.	The utility of AI-prompted re-reflection required active human monitoring; unguided interactions occasionally led to surface-level reflections and uncritical acceptance of biased feedback.	Qualitative Action Research; MMAT: 100% (5/5)	RQ4
Powell & Courchesne (2024)	USA; N=2 candidate lesson-planning artifacts modeled and evaluated in an exploratory case study.	ChatGPT (GPT-3.5/4); 30-minute prompt-and-response iteration cycles utilizing a 5E instructional structure.	Rapid prototyping produced a complete, structurally coherent lesson plan aligned with state frameworks within 30 minutes, showing high efficiency in early ideation.	The generated lesson plan contained scientifically questionable content representations, omitted critical developmental details, and cited a completely fabricated educational book.	Qualitative Exploratory Case Study; MMAT: 80% (4/5)	RQ1

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
Yorulmaz et al. (2025)	Turkey; N=46 pre-service primary teachers split into equal experimental (N=23) and control (N=23) cohorts.	ChatGPT; comparative analysis of STEM lesson plans designed with and without conversational AI assistance.	The experimental group designed significantly higher-quality STEM plans, demonstrating superior skills in integrating multiple disciplines and designing interactive environments.	Despite quantitative improvements, candidates in the experimental group struggled with precise prompting, requiring active tutor feedback to resolve simplistic instructional ideas.	Quantitative Quasi-Experimental; MMAT: 100% (5/5)	RQ1
Yu & Li (2026)	USA; N=50 early childhood education (ECE) teacher candidates enrolled across two universities.	ChatGPT and DALL-E; action research implementing a customized, culturally responsive AI curriculum module.	Trainees demonstrated significant growth in professional AI literacy, learning to modify algorithmic directions and build culturally responsive lessons for diverse learners.	Initial automated lesson outputs consistently exhibited Western-centric biases and lacked localized cultural nuances, necessitating intense critical adaptation by the candidates.	Qualitative Action Research; MMAT: 100% (5/5)	RQ1
Yang et al., (2026)	USA; N=2 graduate-level pre-service early childhood candidates undergoing active coursework.	DALL-E 3, Imagine Art AI, NotebookLM, and SunoAI; digital name-storytelling projects.	Sustained, repeated trials built prompt refinement strategies, allowing candidates to successfully utilize multimodal generation to support early-years creative expression.	Candidates experienced significant operational anxiety regarding unpredictable media outputs and struggled to maintain instructional agency over automated content.	Qualitative Multiple Case Study; MMAT: 80% (4/5)	RQ1
Lazzeretti (2026)	Italy; N=75 fifth-year pre-service primary teachers working in 23 collaborative groups.	ChatGPT and multimodal generators; analysis of teaching units and detailed AI usage reports.	Trainees used AI effectively as a cognitive catalyst for generating visuals, activities, and revising text, while retaining primary ownership of core instructional goals and sequencing.	Prompting behaviors were highly underspecified; candidates frequently struggled to detect lexical overload and developmentally inappropriate language in AI-generated English materials.	Mixed-Methods Case Study; MMAT: 100% (5/5)	RQ1
Jiang & Li (2026)	China; N=259 pre-service primary teachers evaluated in a quasi-experimental design.	Curated Generative AI tools integrated into a structured triadic instructional model.	The experimental group achieved significantly higher gains across all Intelligent-TPACK dimensions, reduced perceived technology integration pressure, and	Hands-on, structured tasks were mandatory; without continuous programmatic support, candidates' self-directed use reverted to passive, superficial content	Quantitative Quasi-Experimental; MMAT: 100% (5/5)	RQ1

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
			designed superior lessons.	extraction.		
Tran et al., (2025)	Vietnam and Finland; N=33 pre-service teachers split into mindset-with-tools vs. tools-only groups.	ChatGPT; process-mining analysis of screen-recorded interaction events during planning.	Mindset-trained candidates utilized "Reflective Iterative" prompting, showing superior shared regulation and scoring higher on TPACK-informed lesson plan rubrics.	Tools-only candidates demonstrated "Linear Extraction," extracting finalized text with minimal revision, showing blind trust in inaccurate outputs and bypassing pedagogical reasoning.	Mixed-Methods Process Mining; MMAT: 100% (5/5)	RQ4
Gold et al., (2026)	USA; N=70 pre-service PK-Grade 5 and intervention specialist candidates.	Cross-disciplinary GenAI seminar co-taught by methods professors utilizing ChatGPT and educational AI portals.	Trainees shifted from conceptualizing AI as a simple planning engine to utilizing it for complex differentiation, diagnostic assessments, and professional family communication.	Seminar analysis revealed early uncritical copying of AI materials; candidates struggled to align standardized outputs with the specific developmental needs of early childhood classes.	Qualitative Case Study; MMAT: 100% (5/5)	RQ1
Lee (2026)	South Korea; N=12 lesson plans designed by 6 collaborative groups in a science methods course.	Generative AI text and digital modeling tools; qualitative content analysis of drafts and revisions.	Candidate groups revised initial drafts to move AI outputs from basic content presentation to the consolidation stage, significantly increasing the explicitness of scientific concepts.	Initial drafts embedded core concepts too superficially within engagement activities, demonstrating a persistent pedagogical gap in positioning AI for conceptual generalization.	Qualitative Content Analysis; MMAT: 80% (4/5)	RQ1
Karlin et al., (2026)	USA; N=17 survey respondents, 8 interviews, and 2 lesson design cases of elementary candidates.	ChatGPT; qualitative analysis of survey, interviews, and design challenge artifacts.	Candidates possessed highly positive views of AI's workload-reduction capacities but expressed severe ethical concerns regarding the erosion of authentic teacher-pupil relationships.	Candidates with strong subject-matter content knowledge easily identified clear pedagogical limits and informational shortcomings in ChatGPT's structured lesson outputs.	Mixed-Methods Case Study; MMAT: 100% (5/5)	RQ3
Rhodes et al. (2026)	England; N=50+ primary PGCE trainees, university	TeachMateAI and ChatGPT;	Most trainees strategically integrated AI to manage	Trainees reported that AI lesson plan timings were	Qualitative Multiple Case	RQ4

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
	tutors, and school placement mentors.	longitudinal, multi-stakeholder school placement partnership study.	workload-intensive, repetitive design tasks (scaffolding worksheets, model texts) rather than copying complete lessons.	highly unrealistic, generated content was overly repetitive, and specific age-level pitch was frequently misaligned with classroom contexts.	Study; MMAT: 100% (5/5)	
Ku (2026)	South Korea; N=11 pre-service elementary teachers undergoing a required methods course.	Prompt engineering protocols applied to build custom AI chatbots acting as virtual students and mentors.	Trainees engaged in instrumental genesis, designing chatbots that scaffolded conceptual understanding, guided problem-posing, and provided contextual real-world examples.	Candidates faced ongoing frustrations in reconciling their specific pedagogical intentions with the unpredictable prompt interpretation and curriculum fidelity limits of LLMs.	Qualitative Case Study; MMAT: 100% (5/5)	RQ2
Waterfield (2025)	USA; N=15 pre-service elementary and special education candidates.	ChatGPT; mixed-methods comparison of IEP goals written with and without large language models.	AI-assisted workflows significantly expedited compliance-oriented writing and goal structuring, preserving candidate time for active pedagogical engagement.	Automated goals frequently defaulted to generic structures that failed to address highly specific, student-centered diagnostic conditions and disability accommodations.	Mixed-Methods Comparative Study; MMAT: 80% (4/5)	RQ1
Campillo-Ferrer et al. (2025); Creswell (2018)	Spain; N=133 pre-service primary teachers enrolled in teacher preparation programs.	Generative AI portals and deepfake platforms; quantitative analysis of verification behaviors.	Candidates developed deep critical awareness of synthetic media, misinformation risks, and deepfakes, learning to utilize AI constructively for historical inquiry.	Pre-service teachers exhibited an alarming initial vulnerability to deepfake visual sources, highlighting a critical programmatic need for media verification protocols.	Quantitative Experimental; MMAT: 100% (5/5)	RQ3
T. Tusyanah et al. (2026)	Indonesia; N=104 pre-service elementary education candidates.	Diverse Generative AI interfaces (ChatGPT, Google Gemini).	AI readiness significantly associated with innovative lesson design, fully mediating the pathways between technology self-efficacy, perceived usefulness, and pedagogy.	Candidates with low baseline digital self-efficacy experienced high software anxiety and cognitive overload, widening the pedagogical performance gap.	Quantitative Descriptive (SEM); MMAT: 100% (5/5)	RQ4

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
Pai (2026)	USA; N=22 undergraduate dual early childhood/childhood certification candidates.	ChatGPT-4o; customized prompt modules designed for student-thinking analysis.	Candidates practiced analyzing authentic multilingual learner math scripts and evaluated AI's capabilities in generating personalized, scaffolded corrective feedback.	ChatGPT frequently generated overly dense mathematical explanations containing vocabulary unsuitable for young, emergent bilingual learners.	Qualitative Case Study; MMAT: 80% (4/5)	RQ1
Ding et al. (2026)	USA; interdisciplinary computer science and mathematics education project.	AI-driven virtual classrooms using highly trained large language model "student agents".	The simulated "sandbox" allowed candidates to repeatedly practice interactive questioning and manage mathematics discussions in a low-stakes environment, boosting teaching self-efficacy.	Highly dependent on initial system programming; simulated student agents occasionally exhibited erratic conceptual behaviors or failed to replicate the emotional realities of a live classroom.	Qualitative Design-Based Research; MMAT: 100% (5/5)	RQ2
Yu & Mutegi (2026)	USA; collaborative research study.	Multimodal Generative AI tools and conversational chatbots.	Trainees negotiated inclusive teacher identities by using generative AI to produce visual classroom mapping scenarios, exploring diverse instructional layouts and representation formats.	Uncritical reliance on AI configurations occasionally reinforced historical, ableist, or exclusionary classroom zoning designs embedded in the AI's underlying training data.	Qualitative Action Research; MMAT: 100% (5/5)	RQ2
Kang, 2026	China and USA; multi-institutional cohort.	Generative AI assistants integrated into online collaborative peer-argumentation platforms.	Interacting with GenAI during peer-to-peer debates significantly enhanced candidates' prompt precision, argumentation quality, and collaborative peer-regulation behaviors.	Candidates with lower initial AI literacy struggled to formulate logical counter-prompts, experiencing cognitive frustration when the model failed to interpret their instructional rationale.	Qualitative Content Analysis; MMAT: 100% (5/5)	RQ2
Kalenda et al., (2025)	USA; N=59 undergraduate and graduate pre-service candidates.	ChatGPT; guided analytical pre/post evaluation of automated planning scripts.	Candidates' positive perceptions regarding ChatGPT's ability to output finalized, ready-to-enact lessons decreased	Post-analysis reflections highlighted inadequate developmental pacing, unrealistic classroom durations, and absent	Mixed-Methods Evaluative Study; MMAT: 100% (5/5)	RQ3

Author(s) & Year	Country & Sample (N, Level)	GenAI Modality & Workflow	Key Measured Pedagogical Shift	Critical Evaluative Gap / Hallucination Risk	Methodological Design & MMAT Score	Targeted Research Question
			significantly after they completed guided evaluations of the generated content.	differentiated scaffolds for English language learners.		
Kang (2026)	South Korea; empirical evaluation of model performance.	ChatGPT-4, Gemini-2.0 Flash, and DeepSeek-R1.	Models successfully diagnosed conceptual word problem errors but performed poorly on computational errors, highlighting that the educational utility of AI depends entirely on teacher content expertise.	All three models exhibited inconsistent diagnostic patterns; DeepSeek-R1 and ChatGPT-4o frequently generated mathematically invalid justifications for computational errors.	Quantitative Evaluation Study; MMAT: 80% (4/5)	RQ3
Jin et al. (2025)	USA; N=29 pre-service elementary science teachers.	ChatGPT; collaborative Analyzing and Interpreting Student Responses (AISR) tasks.	Trainees iteratively designed, enacted, and revised formative assessment tasks, utilizing ChatGPT to generate rubrics and analyze pupil written work.	ChatGPT frequently failed to conceptualize evolutionary mechanisms accurately, often confusing scientific inquiry processes with basic content acquisition.	Qualitative Case Study; MMAT: 100% (5/5)	RQ3

Shift in Instructional Design & Differentiated Concrete Scaffolding (Addressing Rq1)

The synthesis of these N=25 empirical investigations highlights a fundamental operational shift in how pre-service elementary teachers approach instructional design and material creation. Historically, teacher candidates functioned as passive consumers of rigid commercial textbooks, static digital worksheets, or centralized online lesson plan repositories, experiencing substantial temporal and cognitive burdens when manually designing high-quality, highly differentiated resources. The empirical findings demonstrate that generative AI fundamentally transitions candidates into active, prompt-based developers capable of rapidly engineering customized, concrete-operational learning media tailored to primary classrooms.

In both mathematics and science instruction, where young learners aged five to twelve rely on concrete-operational scaffolding to bridge the abstract and the physical (Piaget, 1964; Bruner, 1966), candidates strategically leverage Large Language Models to translate abstract domain principles into child-accessible pictorial, narrative, and task-based representations. Specifically, pre-service primary educators utilize general LLMs and specialized educational portals such as MagicSchool and Diffit to construct culturally responsive story-based word problems, multi-tiered reading passages adjusted to precise Lexile bands, and highly differentiated visual scaffolding guides, successfully accommodating the heterogeneous readiness levels typical of early childhood classrooms (Gold et al., 2026; Yu & li, 2026).

This structural transition from traditional textbook-bound adaptation to prompt-driven material generation involves distinct cognitive mechanisms. In traditional planning workflows, pre-service teachers allocate substantial extraneous cognitive load to administrative and formatting tasks manually re-typing texts, adjusting font scaling, drawing geometric figures, or formatting worksheets (Sweller et al., 2011). Furthermore, evaluations grounded in self-determination theory indicate that when generative AI functions as a supportive cognitive partner, it enhances pre-service teachers' feelings of instructional competence, fosters their professional autonomy through active content modification, and strengthens peer relatedness during collaborative, dialogic planning cycles (Abualrob, 2025; Kim et al., 2026).

In contrast, candidates who do not receive structured training exhibit high "prompt rigidity" and superficial formatting behaviors, often termed "Linear Extraction" (Tran et al., 2025). These candidates enter short, unstructured prompts and accept the first automated output uncritically, resulting in static, superficial content presentation. In these cases, the generative tool functions merely as a text-refining or administrative mediator, failing to generate deep pedagogical value or address the complex developmental needs of early childhood classes (Gold et al., 2026; Kang, 2026; Kalenda et al., 2025).

Technological Agency Through Prompt-Assisted Code & Interactive Authoring (Addressing Rq2)

A second major thematic finding involves the expansion of pre-service elementary teachers' technological agency and pedagogical independence through prompt-assisted software authoring. Historically, primary educators faced severe technical barriers when seeking to design interactive digital manipulatives or dynamic visual simulations for early numeracy and science instruction. They remained highly dependent on expensive, pre-packaged commercial software ecosystems, proprietary low-code web builders, or static presentation slides, which severely limited their ability to customize digital media to their

specific curricular standards and student demographics.

Recent empirical studies demonstrate that pre-service teachers can completely bypass these technical software development barriers by using natural language prompts to direct advanced generative models in authoring lightweight, interactive code. Candidates with zero formal background in computer science successfully leverage generative conversational models to write functional, bug-free HTML5, CSS, and JavaScript canvas code. This prompt-assisted coding capability democratizes digital resource creation, granting future elementary educators unprecedented technological agency and liberating them from the operational constraints of proprietary software platforms.When the AI-generated code contains logical syntax errors or fails to run correctly in the web browser, candidates with low programming background experience high cognitive frustration and software anxiety (Tusyanah et al., 2026).

Critical Ai Literacy, Hallucination Auditing, and Data Ethics (Addressing Rq3)

The third thematic pillar centers on pre-service elementary teachers' critical AI literacy, focusing specifically on conceptual hallucination verification, algorithmic bias detection, and ethical data stewardship. Across the synthesized corpus, researchers emphasize that functional operational skills in prompt formulation do not automatically translate into critical evaluative judgment.

Empirical assessments reveal a systemic "evaluative vigilance gap" among early-career primary teacher candidates: while candidates express high confidence in using GenAI for resource generation, their ability to systematically audit outputs for subtle domain hallucinations and informational shortcomings remains severely underdeveloped.

In early mathematics lesson planning, LLMs routinely generate word problems characterized by flawed underlying set logic, mathematically invalid justifications, or impossible real-world physical parameters (Kang, 2026; Ku, 2026). For instance, models frequently output fraction multiplication or division word problems where the physical units are non-sensical or violate the mathematical properties of division.

In primary science, conversational agents frequently produce oversimplified explanations that introduce deep-seated scientific misconceptions such as misrepresenting plant photosynthesis, water state changes, stellar constellations, or lunar phases which pre-service teachers with weak subject-matter content knowledge often absorb uncritically (Choi & Kim, 2026; Powell & Courchesne, 2024).

This vulnerability is heavily mediated by the candidate's foundational Content Knowledge (CK). Pre-service teachers with robust subject-matter content knowledge easily identify clear pedagogical limits and informational shortcomings in ChatGPT's lesson outputs (Yadav et al., 2025), whereas those with weak content knowledge fail to detect subtle conceptual errors, leading to the uncritical adoption and live classroom enactment of flawed AI outputs (Choi & Kim, 2026; Tran et al., 2025).. The high linguistic surface fluency and aesthetic sophistication of generative outputs act as a cognitive trap, inducing automation bias and leading candidates to assume that structural cohesion guarantees conceptual accuracy.

However, when structured "epistemic auditing rubrics" and guided pre/post evaluations are integrated into methods courses, candidate perceptions of AI's infallibility shift dramatically. Studies demonstrate that after completing guided evaluative tasks where they must cross-reference AI-generated planning scripts against peer-reviewed academic databases, candidates' blind trust drops significantly, and they demonstrate a dramatic

e-ISSN: 2809-4085
p-ISSN: 2809-8749

increase in evaluative vigilance and critical verification behaviors (Gold et al., 2026; Kalenda et al., 2025; Yorulmaz et al., 2025).

Beyond domain hallucinations, candidates struggle to detect embedded socio-cultural and algorithmic biases in automated lesson and media outputs. Generative systems consistently output majoritarian, Western-centric narratives and visual representations, completely omitting localized cultural nuances, native historical figures, and indigenous science practices (Yu & Li, 2026; Yu & Mutegi, 2026). Without intense critical adaptation and explicit culturally responsive prompt training, candidates risk reinforcing historical, ableist, or exclusionary stereotypes in early childhood classrooms (Yu & Li, 2026; Yu & Mutegi, 2026).

Practicum Dynamics: the Stress-Adoption Reconfiguration Pattern (Addressing Rq4)

The fourth thematic pillar explores the intersection of workload management, clinical placement stress, and year-level maturity differences among pre-service teachers. Designing effective lesson plans is a complex, high-stakes professional practice that demands the simultaneous integration of curriculum standards, diagnostic assessments, behavioral transitions, and developmental differentiation.

Under the chronic pressure of clinical field placements and live student teaching, the introduction of generative AI tools introduces a distinct Stress-Adoption Reconfiguration Pattern. When confronted with high-stakes placement anxiety and intense preparation pressures, teacher candidates undergo a transition: they shift from cautious, reflective, and iterative tool exploration in university coursework to a state of pragmatic, survival-driven reliance during student teaching (Rhodes et al., 2026; Tran et al., 2025). Under intense time constraints, candidates deploy GenAI as a workload-reduction shield to rapidly produce daily primary lesson plans and classroom-facing materials, prioritizing generation velocity over pedagogical auditing (Rhodes et al., 2026; Tran et al., 2025).

During these high-pressure practicum periods, candidates are most likely to reduce their evaluative audit rigor, and uncritical copying of AI-generated content increases. Because automated lesson outputs possess high surface fluency, candidates frequently bypass the essential micro-sequencing analysis and reflective design practices necessary for developing foundational pedagogical reasoning, leading to the uncritical enactment of lesson plans characterized by unrealistic classroom durations, inadequate developmental pacing, and absent differentiated scaffolds (Kalenda et al., 2025; Rhodes et al., 2026).

In contrast, upperclassmen and graduate-level candidates demonstrate a more pragmatic, stress-resilient reliance, strategically utilizing prompt engineering to manage workload-intensive, repetitive design tasks while maintaining greater professional judgment, highlighting the need for developmental, scaffolded AI integration throughout the teacher preparation program (Yang et al., 2026; Tran et al., 2025).

Discussion

The systematic synthesis of the global empirical literature on Generative Artificial Intelligence (GenAI) integration in pre-service elementary teacher education establishes a multi-dimensional, socio-technical landscape. By examining the findings of diverse international cohorts through integrated theoretical lenses, this discussion articulates the major conceptual contributions of our review.

Theoretical Reconceptualization: Intelligent-Tpack and Cognitive Ergonomics in Primary Teacher Preparation

The empirical evidence across the N=25 included investigations necessitates a fundamental reconceptualization of Mishra & Koehler's (2006) Technological Pedagogical Content Knowledge (TPACK) framework. Within contemporary primary teacher education, the emergence of GenAI has catalyzed a paradigm shift toward Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) and Intelligent-TPACK (I-TPACK) (Celik, 2023; Chiu, 2026; Deb & Sofal, 2026). Unlike conventional, deterministic educational technologies that operate as static media for information presentation or linear administrative retrieval, generative systems function as pseudo-epistemic co-designers and dynamic conversational partners (Celik et al., 2026; Ng et al., 2025). Consequently, I-TPACK is not merely an additive technological overlay but a highly integrated, non-deterministic domain of professional expertise. It requires candidate teachers to possess an interdependent understanding of how probabilistic, large language models (LLMs) behave, how their variable and context-sensitive outputs align with specific curriculum standards, and how these outputs must be mediated to accommodate the developmental needs of young learners (Cao et al., 2026; Celik, 2023).

To comprehend the operational dynamics of this theoretical model, this study draws on two related but conceptually distinct frameworks: Sweller's Cognitive Load Theory (CLT) and Mayer's Cognitive Theory of Multimedia Learning (CTML) (Sweller et al., 2011; Mayer, 2014). CLT provides the foundational architecture, positing that working memory has limited capacity and that instructional design must manage three load types intrinsic, extraneous, and germane to optimize learning. CTML builds directly on this foundation but narrows its scope to multimedia-specific processing: it assumes learners process information through two separate channels (visual/pictorial and auditory/verbal), and that effective multimedia design reduces extraneous load in each channel while supporting germane processing across both.

Traditionally, the manual creation of highly differentiated, concrete-operational resources such as leveled decodable text passages, tiered mathematics word problems, and multimodal visual scaffolds represented an immense temporal and cognitive burden for candidate teachers, frequently exhausting their mental resources on administrative and formatting tasks.

However, Cognitive Load Theory dictates that a reduction in extraneous cognitive strain does not automatically translate into superior pedagogical reasoning or enhanced learning outcomes (Sweller et al., 2011). Instead, the cognitive space freed by automated offloading must be actively redirected toward germane cognitive load specifically, the high-leverage professional tasks of systematic verification, conceptual auditing, source triangulation, and prompt refinement (Lazzeretti, 2026; Lee, 2026; Tran et al., 2025).

The synthesis highlights that this cognitive reallocation is highly volatile and entirely contingent upon the candidate's domain-specific Content Knowledge (CK) and critical AI literacy (Kang, 2026; Tran et al., 2025). When pre-service primary educators possess robust, foundational content knowledge, they successfully deploy their available cognitive bandwidth toward rigorous, reflective evaluation, recognizing subtle mathematical anomalies or scientific misconceptions in LLM outputs (Yadav et al., 2025).

Deconstructing the Dialectical Paradoxes of Generative Ai Integration

The core contribution of this systematic synthesis lies in deconstructing two distinct dialectical paradoxes that complicate the uncritical adoption of GenAI in initial teacher preparation: the "Augmented Agency versus Hallucination Vulnerability" Paradox and the

"Preparation Velocity versus Pedagogical Atrophy" Tension.

The Augmented Agency versus Hallucination Vulnerability Paradox deconstructs the structural relationship between candidate self-efficacy, generation velocity, output surface fluency, and evaluative audit rigor during lesson design. When pre-service teachers leverage generative portals to instantly output structurally complete, professionally formatted lesson plans and instructional media, they experience an immediate, positive surge in self-perceived professional competence and pedagogical autonomy (Abualrob, 2025; Tusyanah et al., 2026).

This vulnerability is exceptionally acute in early childhood and primary education, where instruction must align with concrete-operational developmental parameters (Piaget, 1964; Bruner, 1966). While LLMs output highly readable prose, they routinely generate subtle, highly plausible, but mathematically and scientifically incorrect domain hallucinations in early grade concepts (Choi & Kim, 2026; Powell & Courchesne, 2024). In early mathematics, models frequently generate word problems characterized by invalid set logic, impossible physical parameters, or mathematically incorrect visual fraction representations (Kang, 2026; Ku, 2026). In elementary science, conversational agents routinely produce oversimplified or outdated explanations that reinforce children's misconceptions regarding gravity, water state changes, stellar constellations (Choi & Kim, 2026; Jin et al., 2025).

Comparing search and verification behaviors across the corpus suggests a qualitative contrast: studies describing unstructured, self-directed exploration generally reported weaker attention to hallucination detection, whereas studies involving structured, epistemic auditing rubrics and process-oriented "mindset" interventions described more active evaluative vigilance, including cross-referencing AI-generated claims with authoritative scholarly sources (Tai et al., 2018; Kalenda et al., 2025; Tran et al., 2025).

Designing highly customized primary lesson plans is a complex, high-stakes professional practice that demands the simultaneous coordination of curriculum standards, diagnostic assessments, behavioral transitions, and multi-level developmental differentiation (Kuzu et al., 2025; Lazzeretti, 2026). In methods coursework, where grading criteria prioritize reflective deliberation and iterative design, candidates actively engage with GenAI as a reflective partner, using prompt engineering to explore alternative pedagogical strategies (Abualrob, 2025; Lee, 2026).

This survival-driven reliance risks bypassing the very developmental micro-sequencing and reflective lesson-planning analysis essential for developing foundational pedagogical reasoning, leading to the gradual atrophy of candidate teachers' independent planning skills (Rhodes et al., 2026)

Software Authoring Agency: Reconfiguring Pedagogical Sovereignty and Its Computational Limits

A highly innovative theme synthesized across the corpus is the expansion of pre-service elementary teachers' technological agency and professional autonomy through prompt-assisted software authoring. Historically, primary educators faced severe technological and economic barriers when seeking to design interactive digital manipulatives or dynamic visual simulations tailored to early numeracy and science inquiry (Bruner, 1966).

The empirical evidence demonstrates that generative AI enables candidate teachers with zero formal background in computer science to completely bypass these technical

software development barriers (Creely, 2026). By using natural language prompts to direct advanced large language models, pre-service primary educators successfully author functional, bug-free HTML5, CSS, and JavaScript canvas code (Alanazi, 2019; Creely, 2026).

The Concrete-Pictorial-Abstract (CPA) framework operationalizes Bruner's (1966) tripartite modes of representation enactive, iconic, and symbolic into an instructional progression widely recognized in primary mathematics pedagogy (Leong et al., 2015). This prompt-assisted coding capability democratizes educational software development, granting future primary educators unprecedented technological agency and liberating them from the operational constraints of proprietary platforms.

However, the synthesis also uncovers a critical technical bottleneck within this emergent software authoring workflow. While candidates can rapidly generate initial functional code, they frequently encounter severe cognitive bottlenecks when attempting to execute complex iterations or debug programmatic errors (Creely, 2026). Because pre-service teachers lack foundational training in computer science, software architecture, and debugging logic, they struggle to formulate logical counter-prompts when code fails to compile or exhibits unexpected visual behaviors.

This operational limitation triggers significant software anxiety, technical apprehension, and cognitive frustration, occasionally causing candidates to abandon the interactive manipulative design entirely and revert to static, pre-packaged media (Creely, 2026). Consequently, the professional viability of prompt-assisted software authoring is strictly bounded by the candidate's baseline digital literacy and the availability of structured, programmatic scaffolding within their methods courses (Kim et al., 2026; Tusyanah et al., 2026).

Critical Evaluative Judgment, Algorithmic Bias, and Child Data Ethics

Beyond cognitive and operational dimensions, the integration of generative AI in elementary teacher preparation exposes candidate teachers to severe ethical and regulatory vulnerabilities. Chief among these is the critical breakdown between candidates' operational self-efficacy and their actual regulatory compliance regarding pupil data protection (Goldstein et al., 2026; Waterfield, 2025). Under intense workload pressures during student teaching, candidates frequently input un-anonymized primary student diagnostic data, behavioral portfolios, and special educational needs profiles directly into public generative endpoints to rapidly generate customized lesson plans or Individualized Education Program (IEP) goals (Chiu & Sanusi, 2024; Waterfield, 2025).

This ethical vulnerability is further compounded by the uncritical adoption of Western-centric algorithmic representations in early childhood storytelling and visual media. While structured, culturally responsive AI curriculum modules successfully build pre-service teachers' capacity to modify algorithmic directions and adapt visual media to reflect diverse student identities, the synthesis demonstrates that such critical mediation is not an automatic behavior (Yu & Li, 2026). Without explicit, programmatic instruction in algorithmic bias auditing and culturally responsive prompting, candidates remain highly susceptible to reproducing exclusionary pedagogical designs in diverse early childhood classrooms (Cain, 2024).

Conclusion

This systematic evidence synthesis evaluated the global literature across N=25 studies to map the pedagogical, cognitive, and structural dynamics of Generative Artificial

Intelligence (GenAI) integration in pre-service elementary and early childhood teacher preparation. The synthesized evidence reveals a transition from passive digital media consumption to active, prompt-driven pedagogical co-design, where candidate teachers leverage large language models to construct differentiated Concrete-Pictorial-Abstract (CPA) scaffolding suitable for concrete-operational learners aged 5–12. Methodologically, the current body of synthesized evidence is constrained by a predominance of short-term intervention horizons and self-reported competencies that lack direct, classroom-enacted validation. Additionally, the literature exhibits a critical gap regarding longitudinal transfer data, as none of the analyzed studies systematically measure the direct downstream impact of GenAI-designed materials on K–5 student learning outcomes, socio-emotional development, or conceptual progression. To resolve these limitations, future research must shift from immediate self-efficacy measures toward longitudinal field trials that evaluate candidate performance during extended teaching placements using structured epistemic auditing rubrics and process-oriented logs. Finally, teacher preparation curricula must establish strict, programmatic child data governance protocols to mitigate the risk of candidate teachers inputting un-anonymized primary pupil diagnostic data into public generative endpoints, thereby ensuring secure, ethical, and developmentally vigilant classroom practices.

References

Abualrob, M. (2025). From reflection to re-reflection: ChatGPT and transformative learning in preservice elementary science education. *International Journal of Learning, Teaching and Educational Research*, 24(10), 769–793. https://doi.org/10.26803/ijlter.24.10.37

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Alanazi, M. H. (2019). A study of the pre-service trainee teachers problems in designing lesson plans. *Arab World English Journal*, 10(1), 166–182. https://doi.org/10.24093/awej/vol10no1.15

Bae, H., Hur, J., Park, J., Choi, G. W., & Moon, J. (2024). Pre-service teachers' dual perspectives on generative AI: Benefits, challenges, and integrating into teaching and learning. *Online Learning*, 28(3), 131–156. https://doi.org/10.24059/olj.v28i3.4543

Braun, V., & Clarke, V. (2023). Thematic analysis. In H. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. Rindskopf, & S. J. Sher (Eds.), *APA handbook of research methods in psychology: Research designs: Quantitative, qualitative, neuropsychological, and biological* (2nd ed., pp. 65–81). American Psychological Association. https://doi.org/10.1037/0000319-004

Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

Campillo-Ferrer, J. M., López-García, A., & Miralles-Sánchez, P. (2025). Student perceptions of the use of Gen-AI in a higher education program in Spain. *Digital*, 5(3), 29. https://doi.org/10.3390/digital5030029

Cao, X., Huang, Z., Li, M., & He, T. (2026). Teachers' AI-TPACK as a tangible outcome in the digital transformation of education: A machine learning-based multilevel approach.

Teaching and Teacher Education, 169, 105270. https://doi.org/10.1016/j.tate.2025.105270

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, 107468. https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Kontkanen, S., Laru, J., & Dalyanci, A. A. (2026). Co-constructing adaptive lesson plans with GenAI: Pre-service teachers' Intelligent-TPACK and prompt engineering strategies. *Computers & Education*, 241, 105485. https://doi.org/10.1016/j.compedu.2025.105485

Cerveró-Carrascosa, A., & Di Sarno-García, S. (2025). DeepSeek's feedback on short essay-writing in EFL: Pre-service primary school teachers' perceptions. *The EuroCALL Review*, 32(2), 38–48. https://doi.org/10.4995/eurocall.2025.23951

Chiu, T. K. F. (2026). Intelligent-TPACK (I-TPACK) framework developed from TPACK through integration of artificial intelligence literacy and competency. *Interactive Learning Environments*, 34(10), 1–16. https://doi.org/10.1080/10494820.2026.2615818

Chiu, T. K. F., & Sanusi, I. T. (2024). Define, foster, and assess student and teacher AI literacy and competency for all: Current status and future research direction. *Computers and Education Open*, 7, 100182. https://doi.org/10.1016/j.caeo.2024.100182

Choi, Y.-S., & Kim, M.-C. (2026). Exploring elementary pre-service teachers' AI literacy in ChatGPT-assisted science lesson design. *Education Sciences*, 16(8), 1275. https://doi.org/10.3390/educsci16081275

Coleman, O. F., & Waterfield, D. A. (2026). Ethical AI use in IEP development: A guiding framework. *Journal of Special Education Technology*, 41(1), 1–15. https://doi.org/10.1177/01626434261419099

Creely, E. (2026). Generative AI to foster computational thinking in initial teacher education: A thematic literature review and model. *Preprints.org*, 2026021017.v1. https://doi.org/10.20944/preprints202602.1017.v1

Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Critical Appraisal Skills Programme. (2018). *CASP qualitative checklist*. CASP.

Deb, J. P., & Sofal, F. A. (2026). TPACK to I-TPACK: A systematic analysis of I-TPACK (integrating intelligence or artificial intelligence with technological pedagogical content knowledge). *International Journal of Emerging Voices in Education*, 2(1), 1–11. https://doi.org/10.59828/ijeve.v2i1.26

Ding, M., Fukawa-Connelly, T., & Liu, D. (2026, February 11). *Turning interdisciplinary collaboration into future vision for education*. Temple University College of Education and Human Development News. https://education.temple.edu/news/2026/02/turning-interdisciplinary-collaboration-future-vision-education

Dirgantara, T., Triayomi, R., Murwanto, P., & Kurniawati, A. F. (2026). Analysis of generative AI applications usage by elementary school teacher education students in supporting the learning process. *Jurnal Pendidikan Rafflesia*, 4(2), 149–158.

Durmus, C. (2024). From AI-generated lesson plans to the real-life classes: Explored by pre-service teachers. *Proceedings of the 10th International Conference on Higher Education Advances (HEAd'24)*, 985–992. Editorial Universitat Politècnica de

València. https://doi.org/10.4995/head24.2024.17060

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255–284. https://doi.org/10.1080/15391523.2010.10782551

Flavin, E., & Kang, Y. (2025). Primary mathematics methods: Diagnosing and remediating student errors in fraction multiplication using ChatGPT-4, Gemini-2.0 Flash, and DeepSeek-R1. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(12), Article em2742. https://doi.org/10.29333/ejmste/17178

Frøsig, J., & Romero, M. (2024). Reconceptualizing teacher agency in human-AI collaborative design. *Journal of Digital Learning in Teacher Education*, 40(4), 185–199. https://doi.org/10.1080/21532974.2024.2367573

Gold, L. A., Winn, V., & Arnold, J. M. (2026). Exploring the impact of generative AI on curriculum design and instruction: A study with preservice teachers. *Thomas C. Hunt Building a Research Community Day*, 75. https://ecommons.udayton.edu/sehs_brc/75

Goldstein, O., Marae-Haj, N., & Zidan, W. (2026). Pioneering teacher educators navigating AI integration in pre-service teacher preparation: Strategies and challenges. *AI Education*, 2(3), 29. https://doi.org/10.3390/aieduc2030029

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or rapid reviews? Evaluating the retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181–217. https://doi.org/10.1002/jrsm.1378

Hong, Q. N., Pluye, P., Fàbregues, S., Regnault, L., Brière, N., Granikov, V., Boardman, F., Ting, L., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Registration of Copyright, Canadian Intellectual Property Office.

Hossain, S., Ahmadi, S., Li, L., Awoyemi, I. D., Huang, W., Zhou, C., Li, J., Haniya, S., Khanam, S., & Subaha, T. M. (2026). *Rethinking generative AI literacy: An integrative, developmental, and dialectical framework for K–12 teacher education*. arXiv. https://doi.org/10.48550/arXiv.2608.01705

Hsu, H.-P. (2026). Artificial intelligence literacy and competency in pre-service teacher education. *Encyclopedia*, 6(4), 76. https://doi.org/10.3390/encyclopedia6040076

Jiang, S., & Li, J. (2026). Triadic instructional design: The impact of structured AI training on pre-service teachers' Intelligent-TPACK, attitudes, and lesson planning skills. *Education Sciences*, 16(2), 315. https://doi.org/10.3390/educsci16020315

Jin, H., Cisterna, D., & Riordan, B. (2025). Pre-service elementary science teachers analyzing and interpreting student responses on natural selection using ChatGPT. *Journal of Science Teacher Education*, 36(6), 654–672. https://doi.org/10.1186/s43031-025-00146-8

Kalenda, P. J., Rath, L., Abugasea Heidt, M., & Wright, A. (2025). Pre-service Teacher Perceptions of ChatGPT for Lesson Plan Generation. *Journal of Educational Technology Systems*, *53*(3), 219-241. https://doi.org/10.1177/00472395241301388

Kang, H. M. (2026). Exploring pre-service elementary teachers' prompting as shaping practices when using ChatGPT for fraction division lesson planning. *The Journal of Educational Research in Mathematics*, 36(2), 233–252. https://doi.org/10.29275/jerm.2026.36.2.233

Karlin, M., Stephany, C., Okamura, K., & Nam-Huh, Y. J. (2026). Balancing beliefs: Exploring preservice teachers' generative artificial intelligence perceptions and intentions for practice. *Frontiers in Artificial Intelligence*, 9, Article 1834667. https://doi.org/10.3389/frai.2026.1834667

Kim, J., Yu, S., Detrick, R., Fan, L., & Li, N. (2026). Students' perception of generative AI-assisted collaborative argumentation. *European Journal of Education*, 61(2), Article e70576. https://doi.org/10.1111/ejed.70576

Korucu-Kış, S., & Bakla, A. (2026). Exploring pre-service EFL teachers' perceptions, use cases and questioning practices of ChatGPT during practicum: A critical digital pedagogy perspective. *Instructional Science*, 54, Article 33. https://doi.org/10.1007/s11251-026-09783-6

Ku, N. (2026). Exploring elementary pre-service teachers' instrumental genesis in designing AI chatbots. *Eurasia Journal of Mathematics, Science and Technology Education*, 22(3), Article em2792. https://doi.org/10.29333/ejmste/18067

Kuzu, T. E., Irion, T., & Bay, W. (2025). AI-based task development in teacher education: An empirical study on using ChatGPT to create complex multilingual tasks in the context of primary education. *Education and Information Technologies*, 30, 23041–23075. https://doi.org/10.1007/s10639-025-13673-8

Lazzeretti, C. (2026). AI-supported design of teaching units for English to young learners: A case study in initial teacher education. *Education Sciences*, 16(4), 614. https://doi.org/10.3390/educsci16040614

Lee, J. (2026). How pre-service elementary teachers develop scientific concepts in AI-integrated lesson designs: Implications for sustainable teacher education. *Sustainability*, 18(10), 5211. https://doi.org/10.3390/su18105211

Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, 55(5), 1982–2002. https://doi.org/10.1111/bjet.13505

Leong, K. E., Eow, Y. L., & Rahim, S. S. A. (2015). Secondary mathematics teachers' technological pedagogical content knowledge (TPACK) and their instructional practices. *The Asia-Pacific Education Researcher*, 24(3), 441–450. https://doi.org/10.1007/s40299-014-0195-2

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. https://doi.org/10.11613/BM.2012.031

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Onbaşılı, Ü. I. (2026). Empowering pre-service teachers with generative AI: A GenAI-TPACK-based approach to digital storytelling. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-026-13991-5

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw,

J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. https://doi.org/10.1136/bmj.n71

Pai, G. (2026). *Using AI to enhance math teaching for multilingual learners: An asynchronous interactive scenario*. AmplifyLearn.AI Use Case, University of Washington & Queens College, City University of New York. https://www.amplifylearn.ai/grace-pai-use-case-2/

Piaget, J. (1964). *Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), Piaget rediscovered: A report on the conference of cognitive studies and curriculum development* (pp. 7–20). Cornell University.

Powell, W., & Courchesne, S. (2024). Opportunities and risks involved in using ChatGPT to create first grade science lesson plans. *PLoS ONE*, 19(6), Article e0305337. https://doi.org/10.1371/journal.pone.0305337

Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury Publishing.

Rhodes, A., Birchinall, E., Kilkenny, K., James, D., & Jayson, N. (2026). *Embracing generative AI in initial teacher education: Research, impact and innovation* [Presentation]. Figshare, University of Manchester. https://doi.org/10.6084/m9.figshare.26050428

Rind, I. A. (2026). Conceptualizing the impact of AI on teacher knowledge and expertise: A cognitive load perspective. *Education Sciences*, 16(1), Article 57. https://doi.org/10.3390/educsci16010057

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. https://doi.org/10.3102/0013189X015002004

Sun, L., & Bakla, A. (2026). The predictive role of frequency of use on teachers' generative AI ethical reflection: Scale development and validation. *Journal of Research on Technology in Education*, 58(3), 415–430. https://doi.org/10.1080/15391523.2026.2711656

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-8126-4

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students' capability for future learning. *Higher Education*, 76(3), 467–481. https://doi.org/10.1007/s10734-017-0220-z

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, Article 45. https://doi.org/10.1186/1471-2288-8-45

Tran, P. T. H., Huynh, L., Bien, T.-A., Dang, B., & Nguyen, A. (2025). Preparing preservice teachers for generative AI in lesson planning: A process mining study of AI mindset and tool-only training. *Journal of Digital Learning in Teacher Education*, 42(1), 16–32. https://doi.org/10.1080/21532974.2025.2583516

Tusyanah, T., Samsudin, N., Ismiyati, I., Chayati, N., Pramusinto, H., & Suryanto, E. (2026). The pathway from readiness to pedagogy: AI readiness as a catalyst for innovative lesson-designing in preservice teachers. *Journal of Pedagogical Sociology and Psychology*, 8(2), Article e42865. https://doi.org/10.33902/JPSP.202642865

Waterfield, D. A., Coleman, O. F., Welker, N. P., Kennedy, M. J., McDonald, S. D., & Cook, B. G. (2025). IEPs in the age of AI: Examining IEP goals written with and without ChatGPT.

Journal of Special Education Technology, 40(4), 215–231. https://doi.org/10.1177/01626434251324592

Yadav, A., Lachney, M., Hu, A., & Tavernier, L. (2025). Integrating generative AI in teacher education: Pre-service teachers' perspectives, attitudes, and design challenges. *Journal of Research in Childhood Education*, 40(1), 24–45. https://doi.org/10.1080/02568543.2025.2581712

Yang, S., Arbor, A., & Chubin, S. (2026). *AI-powered digital storytelling: Preservice teachers' perceptions of GenAI in early childhood education*. University of Maryland, Baltimore County (MDSOAR). https://mdsoar.org/handle/11603/34415

Yorulmaz, A., Okulu, H. Z., Muslu-Komurcu, N., & Cokcaliskan, H. (2025). Enhancing STEM lesson plans: preservice primary teachers' collaboration with ChatGPT. *Education and Information Technologies*, *30*(17), 24543-24573. https://doi.org/10.1007/s10639-025-13736-w

Yu, S., & Li, X. (2026). AI literacy and culturally responsive lesson design in early childhood teacher education. *Teaching Education*. Advance online publication. https://doi.org/10.1080/10476210.2026.2642691

Yu, S., & Libby-Reynolds, J. (2026). Becoming-with inclusivity in early childhood education professionalism through multimodal visualization of classrooms. *Contemporary Issues in Early Childhood*. Advance online publication. https://doi.org/10.1177/14639491261446062

Yu, S., & Smith-Mutegi, D. (2026). Becoming an inclusive teacher across 'virtual and real' contexts-negotiating teacher identity through engagement with GenAI. *Pedagogy, Culture & Society*, 1–23. https://doi.org/10.1080/14681366.2026.2668080

Zhang, W., Mita, R., & Darko, E. N. K. O. (2026). Preparing teachers for the age of artificial intelligence: Understanding the challenges and needs of AI-TPACK in Indonesian elementary education. *Frontiers in Education*, 11, Article 1769204. https://doi.org/10.3389/feduc.2026.1769204