

K-12 AI literacy frameworks after generative AI: A scoping review

Med Kharbach ^{1*} , Johanathan Woodworth ¹ , Sandra Sawchuk ¹ 

¹Mount Saint Vincent University, Halifax, NS, CANADA

*Corresponding Author: med.kharbach@msvu.ca

Citation: Kharbach, M., Woodworth, J., & Sawchuk, S. (2026). K-12 AI literacy frameworks after generative AI: A scoping review. *International Journal of Professional Development, Learners and Learning*, 8(2), Article e2618. <https://doi.org/10.30935/ijpdll/19665>

ABSTRACT

Artificial intelligence (AI) literacy has become a priority in K-12 education, yet frameworks for teaching and assessing it remain fragmented across grade levels, national contexts, and disciplinary traditions. This scoping review maps and synthesizes framework-level contributions to K-12 AI literacy published between January 2024 and December 2025. Following JBI guidance and PRISMA-ScR reporting standards, we searched four databases and grey literature sources, screened 204 unique records, and applied a purpose-built checklist covering five criteria: grade bands, competencies, critical constructs, assessment, and exemplars. Fifty studies met the inclusion threshold. The corpus spans 15 countries, with the United States (32%), Hong Kong (26%), and mainland China (12%) accounting for 70% of included studies. Frameworks defined AI literacy through cognitive, affective, behavioral, ethical, and social dimensions, but dimensional models used inconsistent labels and constructs, with limited crosswalk among them. The ABCE framework offered the closest approximation to a shared measurement tradition, with adaptation across three educational systems. Assessment and curriculum also developed largely in parallel: scale-development papers reported psychometric evidence without instructional sensitivity data, while curriculum papers assessed learning through lesson-linked tools that limited cross-study comparison. Generative AI competencies, including output evaluation, hallucination detection, and prompt design, remained peripheral in most frameworks despite their relevance to students' everyday encounters with AI. Geographic concentration and limited attention to socioeconomic status, disability, and linguistic diversity further constrained transferability. The field needs stronger convergence through shared vocabulary, curriculum-aligned assessment, longitudinal evidence, and cross-context validation.

Keywords: AI literacy, K-12 education, scoping review, competency frameworks, AI assessment

Received: 02 Jun. 2026 ♦ Accepted: 03 Sep. 2026

INTRODUCTION

Artificial intelligence (AI) has become a defining feature of contemporary society, shaping communication, work, and education (Holmes et al., 2019; Luckin et al., 2016). Children and adolescents now encounter AI through digital assistants, recommendation systems, adaptive learning platforms, and generative models (Ng et al., 2021; Touretzky et al., 2023). In K-12 schooling, AI is increasingly visible not only as a tool for support, but also as a system that can generate text, code, translations, explanations, and other learning artefacts (Su et al., 2023; Zhou et al., 2025). This shift raises educational questions that go beyond access or tool use. It touches learning, authorship, assessment, evidence, and the kinds of judgement students need when AI systems participate in academic work (Knox, 2020; Williamson & Eynon, 2020).

The rapid adoption of AI in schools brings both opportunities and risks. AI-powered platforms can provide adaptive practice, real-time feedback, and personalized support (Zawacki-Richter et al., 2019). At the same time, reliance on automated systems may encourage cognitive offloading, especially when students use AI to bypass memory, reasoning, or problem-solving processes without appropriate scaffolding (Risko & Gilbert, 2016). These concerns are especially important in K-12 settings, where students' executive function, self-regulation, and metacognitive capacities are still developing (Sodian et al., 2020).

Other risks are equally important. Automation bias can lead learners to over-trust AI outputs, even when those outputs are incomplete, misleading, or wrong (Parasuraman & Riley, 1997). Algorithmic systems may also reproduce biases embedded in training data, with consequences for underrepresented groups, languages, and communities

(Baker & Hawn, 2022; UNICEF, 2021). Privacy concerns add another layer, since AI systems often depend on student data, including performance records, behavioral patterns, and communication traces (Holmes et al., 2019; Livingstone et al., 2020). These issues make clear that students need more than functional familiarity with AI tools.

AI literacy addresses this need. It is generally understood as the capacity to understand how AI systems work, use them purposefully, evaluate their outputs and limitations, and consider their ethical and societal implications (Long & Magerko, 2020; Ng et al., 2021). In K-12 contexts, AI literacy also needs to be developmentally appropriate and scaffolded across grade levels. Younger learners do not need the same technical depth as university students or professionals, but they do need structured opportunities to recognize AI, question its outputs, understand its limits, and reason about its effects on people and society.

Despite growing interest, K-12 AI literacy research remains fragmented. Existing studies vary by grade level, national context, disciplinary focus, and theoretical foundation (Su et al., 2023; Zhou et al., 2025). Some frameworks emphasize technical knowledge or programming, others focus on ethical reasoning, and still others develop assessment instruments without clear curricular sequences. International initiatives such as AI4K12 provide important anchors through the five big ideas and grade-band guidance (Touretzky et al., 2023), but the relationship between such initiatives and classroom-level frameworks remains uneven.

This fragmentation creates uncertainty for teachers, curriculum designers, and policymakers. Without structured frameworks that specify what students should know and be able to do across grade bands, AI education may remain limited to tool exposure or isolated activities. A stronger synthesis is needed to distinguish framework-level contributions from isolated activities, tool-use studies, and general discussions of AI in education. It needs to identify which K-12 AI literacy frameworks exist, how they define competencies, how they address critical constructs such as bias and fairness, how they incorporate emerging generative AI (GenAI) competencies, how they assess learning, and what gaps remain. Several recent reviews have mapped AI literacy and AI education from related angles. Zhou et al. (2025) reviewed K-12 AI literacy and competency research, Su et al. (2025) focused on AI literacy assessment, Ma et al. (2025) synthesized K-12 AI ethics education, Liu and Zhong (2024) examined how educators teach AI in K-12 settings, and Tan and Tang (2025) reviewed empirical K-12 AI literacy studies. These reviews provide important coverage of the field, but they do not apply a structured framework-completeness filter to distinguish grade-banded, classroom-ready frameworks from broader AI education studies. The present review addresses this gap through a five-criterion checklist covering grade bands, competencies, critical constructs, assessment, and exemplars. This allows us to map not only what has been

published, but also how complete and usable current K-12 AI literacy frameworks are for curriculum and assessment design.

This scoping review maps, analyses, and synthesizes recent framework-level evidence for K-12 AI literacy published between 2024 and 2025. Five research questions (RQs) guide the inquiry:

- RQ1:** What grade-banded frameworks exist for AI literacy in K-12 education?
- RQ2:** What competencies, critical constructs, and progression models do these frameworks define?
- RQ3:** What assessment strategies are proposed or validated within these frameworks?
- RQ4:** What evidence exists regarding the feasibility and adoption of AI literacy frameworks in schools?
- RQ5:** What limitations and omissions can be identified in current frameworks, particularly regarding assessment, progression, and equity?

METHODS

Study Design

This study is a scoping review, conducted in accordance with JBI methodological guidance for scoping reviews (Aromataris & Munn, 2020; Peters et al., 2020) and reported using the PRISMA-ScR extension (Tricco et al., 2018) and PRISMA-S search reporting guidance (Rethlefsen et al., 2021). The protocol was registered with OSF registries prior to screening.

A scoping review design was selected because the objective was to map the range, structure, and completeness of framework-level contributions to K-12 AI literacy, and to identify where the evidence base is thin. Neither a systematic review nor a meta-analysis would have served this objective. Meta-analysis requires comparable outcome measures drawn from studies sharing a common design, and the eligible corpus comprises conceptual frameworks, scale-validation studies, quasi-experimental interventions, design-based research, and reviews, with no shared outcome metric available for pooling. No effect sizes were combined at any stage of this review.

The design is also methodologically separate from document analysis. Document analysis treats a purposively assembled body of institutional, policy, or organizational documents as qualitative data and interprets them through content or thematic coding. The present study takes published research reports and grey literature as its unit of analysis, identified through a systematic and replicable search of four bibliographic databases and defined grey literature sources, screened independently by two reviewers against predefined eligibility criteria, and charted using a standardized extraction matrix. The procedures reported below follow that sequence.

Eligibility Criteria

Eligible sources were English-language peer-reviewed publications or grey literature documents published between 1 January 2024 and 31 December 2025 that presented, developed, synthesized, or evaluated framework-level K-12 AI literacy contributions. This included competency models, curriculum frameworks, assessment frameworks, pedagogical models, and implementation-oriented frameworks targeting school-aged learners. The 2024-2025 window was selected because it marks a distinct phase in K-12 AI literacy research. The public release of ChatGPT in late 2022 intensified institutional attention to AI in education, but K-12 framework-level responses required time to develop, pilot, and publish. By 2024, post-ChatGPT K-12 frameworks, assessment instruments, and curriculum studies began appearing in peer-reviewed and grey literature sources. Restricting the review to this period allowed us to examine framework-level work produced after GenAI had become a practical classroom concern.

Inclusion required that a paper address at least two of the following five framework criteria:

- (a) grade bands or developmental levels,
- (b) competencies or performance expectations,
- (c) critical constructs such as bias, fairness, accountability, or ethical reasoning,
- (d) an assessment strategy, and
- (e) classroom exemplars or instructional tasks tied to specific levels.

Papers meeting fewer than two criteria were excluded on the grounds that they did not represent a sufficiently structured framework for the purposes of this review.

Also excluded were frameworks targeting adults or higher education exclusively, ethical or philosophical discussions with no competency or assessment component, teacher professional development toolkits that lacked student-facing performance expectations, and single-lesson activities with no broader learning progression. Regional grade-structure models were accepted when they specified equivalent developmental or school-level progression.

Search Strategy

Four databases were searched on 15 January 2026: Academic Search Premier (EBSCOhost), Education Research Complete (EBSCOhost), Education Research Index (ProQuest), and Web of Science Core Collection (Clarivate). The search covered publications from 1 January 2024 to 31 December 2025. Grey literature sources included UNESCO, OECD, EDUCAUSE, Jisc, MIT, ACM SIGCSE, IEEE EDUCON, institutional repositories, and reference lists of included studies.

The search strategy used five conceptual blocks combined with Boolean AND:

- (1) artificial intelligence terminology,

- (2) literacy and competency terminology,
- (3) K-12 educational settings,
- (4) framework and curriculum terminology, and
- (5) human-AI interaction and pedagogy.

Search string development was informed by PRESS guidance (McGowan et al., 2016). The complete search strategies, including platform-specific syntax, are documented in [Supplementary File 1](#).

The four database searches returned 320 records: Academic Search Premier (55), Education Research Complete (70), Education Research Index (105), and Web of Science (90). An additional 19 records were identified through grey literature sources, producing 339 total records.

Study Selection

After removing 135 duplicates, 204 unique records underwent title and abstract screening. Two reviewers independently screened all records, resolving disagreements through discussion. Fifty-four records were excluded at title and abstract screening as clearly outside scope.

The remaining 150 reports were retrieved for full-text assessment against the eligibility criteria. Ninety-five reports were excluded at this stage for reasons including absence of a framework, curriculum, or assessment structure; teacher-focused designs without a student-facing K-12 framework; perception or adoption studies without framework-level contributions; AI use in education without an AI literacy focus; reviews or syntheses without eligible framework content; studies outside K-12 contexts; and wrong publication types. Three duplicate reports that had survived initial deduplication were also identified and removed at this stage. Full-text exclusion categories are reported in [Supplementary File 2](#).

The remaining 55 papers proceeded with framework checklist appraisal. Five papers scored below the 2/5 inclusion threshold and were excluded, producing a final corpus of 50 included studies. Checklist-stage exclusions are summarized in [Supplementary File 2](#).

Framework Checklist Appraisal

Because the purpose of this scoping review was to map framework-level contributions to K-12 AI literacy, conventional methodological quality appraisal was not appropriate. We applied a purpose-built framework checklist to assess the structural completeness of each candidate framework. The checklist comprised the five criteria described above:

- (a) grade bands,
- (b) competencies,
- (c) critical constructs,
- (d) assessment, and
- (e) exemplars.

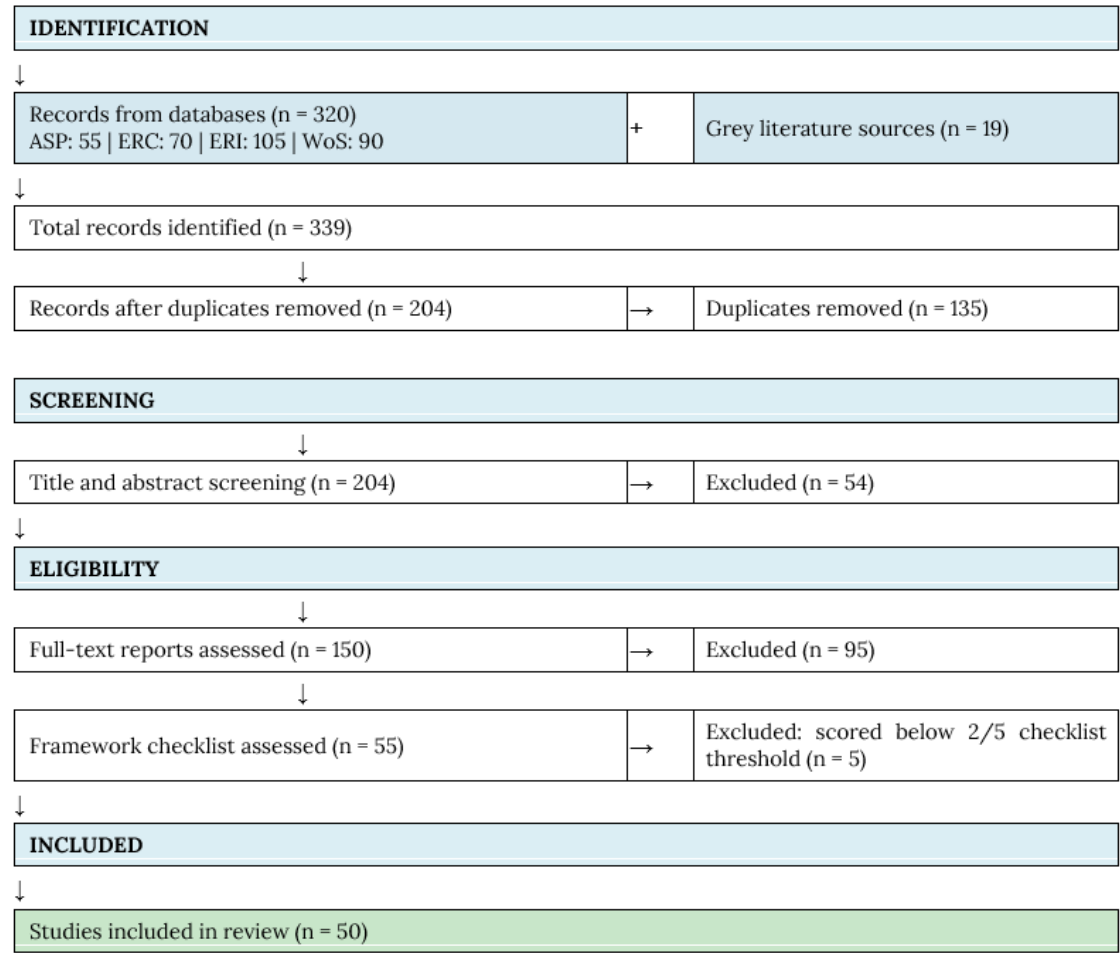


Figure 1. PRISMA-ScR flow diagram of study selection (the authors' study-selection data)

Each criterion was scored as YES (1 point, fully present), PARTIAL (0.5 points, referenced but not sufficiently developed), or NO (0 points, absent). A minimum score of 2 out of 5 was required for inclusion.

Two reviewers independently scored the 55 candidate papers. Disagreements and borderline cases were resolved through discussion until consensus was reached. The scoring rubric, including decision rules and worked examples, is provided in [Supplementary File 3](#). Individual checklist scores for the 50 included studies are reported in the master charting matrix ([Supplementary File 4](#)).

Data Charting

Data charting was conducted using a standardized extraction matrix. For each included study, the following fields were extracted: bibliographic information (authors, year, title, and source), country or region, study type, sample and context, grade-band organization, framework identity and structure, competencies defined, critical constructs addressed, progression model, assessment strategies, feasibility and adoption evidence, stated limitations, and checklist score. Study types and countries were grouped using standardized categories to support reporting in the results and summary tab of the charting matrix. One reviewer charted the data and a second reviewer checked entries, with discrepancies resolved by

consensus. The full charting matrix is provided in [Supplementary File 4](#).

Synthesis

Findings were synthesized narratively and organized by the five RQs. Studies were grouped thematically within each RQ to identify patterns, convergences, and gaps across frameworks. Checklist scores provided a secondary lens for interpreting the structural completeness of individual contributions.

RESULTS

Selection of Sources

As shown in [Figure 1](#), the search identified 339 records, of which 204 remained after duplicate removal. Title and abstract screening excluded 54 records, and full-text assessment excluded 95 reports. Of the 55 candidate papers assessed with the framework checklist, five scored below the 2/5 threshold. The final corpus comprised 50 studies.

Characteristics of Included Studies

The included corpus comprised 50 studies published in 2024 and 2025, the period defined in our eligibility criteria. Thirty-five studies appeared in 2025 and fifteen in 2024.

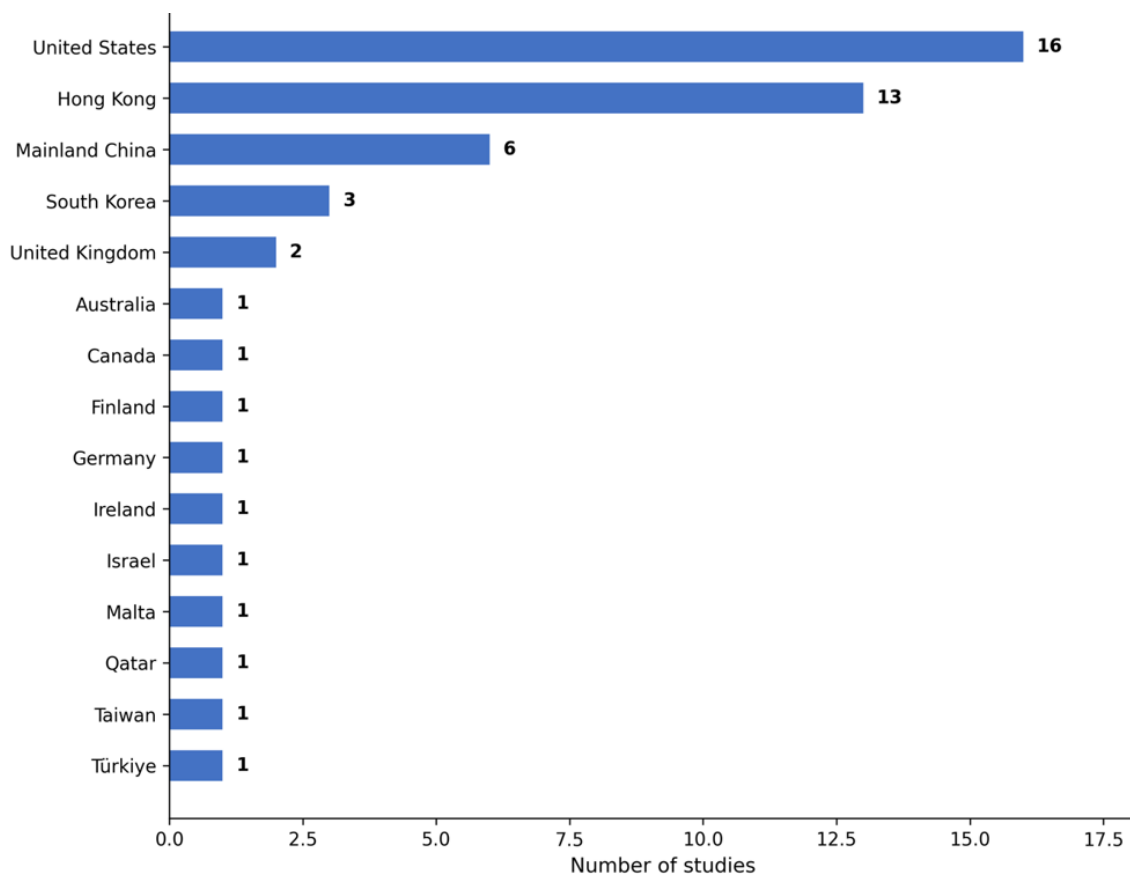


Figure 2. Geographic distribution of included studies (n = 50) (country/region was assigned by primary study setting for empirical studies and by lead author affiliation for review or conceptual papers) (the authors' analysis of included studies)

This concentration shows how quickly framework-level work on K-12 AI literacy has grown in a short period, especially among studies that specify grade bands, competencies, critical constructs, assessment, or classroom exemplars. Earlier AI education work existed, but the present review focused on recent studies with explicit framework-level structure.

The geographic distribution was uneven. The United States contributed the largest share, with sixteen papers, followed by Hong Kong with thirteen and mainland China with six. South Korea contributed three papers, and the United Kingdom contributed two. Ten countries or regions contributed one study each: Australia, Canada, Finland, Germany, Ireland, Israel, Malta, Qatar, Taiwan, and Türkiye. Overall, the United States, Hong Kong, and mainland China accounted for 70% of the corpus, a concentration that limits the generalizability of findings across less represented educational contexts. **Figure 2** summarizes the geographic distribution of the included studies.

Study type was coded by each paper's main contribution. For hybrid papers, we classified the study by its main role in the evidence base: co-design and classroom implementation papers were coded as design-based or curriculum design studies, while framework papers without systematic classroom implementation data were coded as conceptual/framework contributions. Twelve papers used mixed-methods designs. Seven used design-based or curriculum design approaches, and seven were PRISMA-

informed systematic reviews. Five used quasi-experimental or experimental designs, five were narrative or literature reviews, four were qualitative studies, and four focused on scale development and psychometric validation. Two were conceptual/framework papers, two used scoping review methodology, and two used other quantitative approaches, including latent profile analysis and cross-sectional survey design.

Checklist scores also varied. Twenty-one papers scored 5/5, five scored 4.5/5, fourteen scored 4/5, six scored 3/5, and four met the minimum threshold of 2/5. Half-point scores indicated partial coverage, such as mentioning assessment without a concrete strategy or referencing exemplars without implementation detail. Papers scoring 5/5 addressed all five criteria: grade bands, competencies, critical constructs, assessment, and exemplars. These included both measurement-oriented and implementation-rich studies. **Figure 3** visualizes the distribution of checklist scores.

Table 1 summarizes the included corpus by publication year, geographic setting, study type, and framework checklist score; the full study-by-study characteristics table is provided in **Supplementary File 4**.

RQ1. What Grade-Banded AI Literacy Frameworks Exist For K-12 Contexts?

The 50 studies offered varied framework-level contributions, including K-12 curriculum architectures,

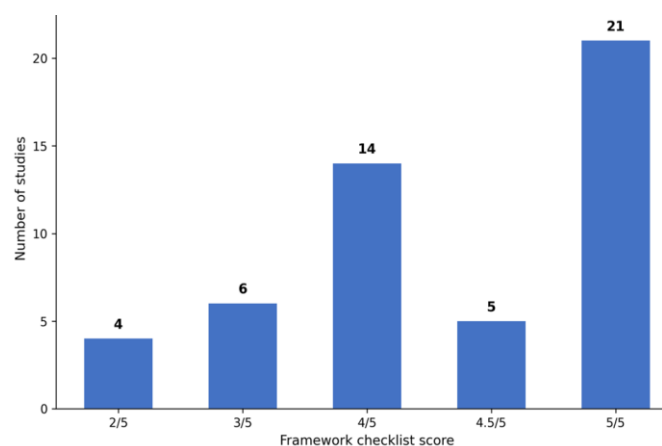


Figure 3. Distribution of included studies by framework checklist score ($n = 50$) (checklist scores range from 2/5 to 5/5 based on five criteria: grade bands, competencies, critical constructs, assessment, and exemplars) (the authors' analysis of the included studies)

single-grade pedagogical models, assessment instruments, and field-mapping reviews. No framework yet provides validated progression, assessment, and classroom exemplars across all grade bands.

Table 2 summarizes four main clusters. Six papers functioned primarily as field-mapping or assessment-mapping reviews: Xia (2025), Lin et al. (2025), Rogers and Carbonaro (2025), Su et al. (2025), Tan and Tang (2025), and Liu and Zhong (2024). These papers inform the **RQ1-RQ5** synthesis by mapping definitions, pedagogical patterns, assessment practices, and research gaps across the wider literature. They are not assigned to a single **RQ1** framework cluster because their main contribution is synthesis, not a discrete grade-banded curriculum framework.

Cross-grade and system-level frameworks

Eleven studies proposed or analyzed frameworks spanning multiple grade bands. Chung et al. (2025) developed a K-12 AI literacy content framework organized around understanding AI concepts, understanding how AI works, and understanding AI ethics, with differentiation across elementary, middle, and high school. Its accompanying AI literacy diagnostic assessment was piloted with 287 sixth-grade students and refined with an additional 293, making it one of the clearest links between a K-12 content framework and grade-level assessment evidence.

Two review-based studies generated cross-grade frameworks. Zhou et al. (2025) separated AI literacy from AI competency and analyzed both across five competency areas and three knowledge domains; their review of 47 studies found no coverage of self-reflection. Ma et al. (2025) reviewed 68 studies across 17 countries and developed a competency-oriented responsible AI literacy framework across kindergarten, elementary, and secondary bands.

Tadimalla and Maher (2024, 2025) proposed an interdisciplinary socio-technical framework covering

Table 1. Summary characteristics of included studies ($n = 50$)

Characteristic	n (%)
Publication year	
2024	15 (30%)
2025	35 (70%)
Country/region of lead author or primary study setting	
United States	16 (32%)
Hong Kong	13 (26%)
Mainland China	6 (12%)
South Korea	3 (6%)
United Kingdom	2 (4%)
Australia	1 (2%)
Canada	1 (2%)
Finland	1 (2%)
Germany	1 (2%)
Ireland	1 (2%)
Israel	1 (2%)
Malta	1 (2%)
Qatar	1 (2%)
Taiwan	1 (2%)
Türkiye	1 (2%)
Study type	
Mixed-methods	12 (24%)
Design-based/curriculum design	7 (14%)
Systematic review	7 (14%)
Narrative/literature review	5 (10%)
Quasi-experimental/experimental	5 (10%)
Qualitative	4 (8%)
Scale development/validation	4 (8%)
Conceptual/framework	2 (4%)
Scoping review	2 (4%)
Other quantitative	2 (4%)
Framework checklist score	
5/5	21 (42%)
4.5/5	5 (10%)
4/5	14 (28%)
3/5	6 (12%)
2/5	4 (8%)

Note. Study types were classified by each paper's primary contribution; see text for coding rationale; Country/region was assigned by primary study setting for empirical studies and by lead author affiliation for review or conceptual papers; Checklist criteria: (a) grade bands, (b) competencies, (c) critical constructs, (d) assessment, (e) exemplars; Score = YES (1) + PARTIAL (0.5) + NO (0); Minimum 2/5 for inclusion. A full study-by-study characteristics table is provided in [Supplementary File 4](#)

technical understanding, GenAI interaction, ethical and responsible use, and societal implications. The 2025 version organized the model into 16 modules, building on the 2024 iteration. Both papers map outcomes against Bloom's revised taxonomy and present the framework as adjustable across learner contexts, though the K-12 evidence base remains limited to higher education teaching experience and a high school summer camp.

Chiu et al. (2024, 2025) contributed two related frameworks. Chiu et al. (2025) developed the student AI competency self-efficacy (SAICS) framework and validated a 32-item, eight-dimension scale through a three-round Delphi process with 29 K-12 AI teachers and confirmatory factor analysis with 448 Hong Kong secondary students. Chiu et al. (2024) proposed a comprehensive framework for K-12 AI education through four co-design cycles with 30 experienced AI teachers from 15 Hong Kong middle schools. The first framework is intended for K-12 but validated with grade 7-grade 12; the second defines five components, technology, impact, ethics, collaboration, and self-

Table 2. Framework clusters identified in the included corpus (n = 44 cluster-assigned papers)

Framework cluster	Main focus	Representative studies	Strength	Main limitation
Cross-grade and system-level (n = 11)	K-12 curriculum architectures, competency matrices, cross-band progression	Chung et al. (2025), Zhou et al. (2025), Ma et al. (2025)	System-level organization; competency mapping across grade bands	Few longitudinal validations; most rely on cross-sectional data or expert consensus
Primary/elementary-focused (n = 6)	AI literacy for young learners (ages 5–12); constructionist and design-based approaches	Yim (2024), Yim and Su (2025), Dai (2024)	Demonstrates feasibility of age-appropriate AI instruction in elementary classrooms	Limited assessment evidence; no tracked developmental progression across primary grades
Secondary-focused competency and assessment (n = 16)	Psychometric instruments, competency scales, and subject-specific frameworks for middle and high school	Ng et al. (2024c), Sun and Chan (2025), Şahin et al. (2025)	Most developed psychometric evidence	Most instruments rely on self-report; few connect to observable classroom performance
Pedagogical, implementation, and design-based (n = 11)	Classroom modules, curriculum prototypes, and iterative design-based research	Masla et al. (2025), Bosch et al. (2025), Smyslova et al. (2025)	Strong classroom-readiness; actionable instructional material with assessment evidence	Mostly pilot or beta-test implementations; short durations; limited evidence of long-term transfer or scalability

Note. Six additional papers classified as field-mapping or assessment-mapping reviews are not assigned to a specific cluster but are cited throughout the RQ1–RQ5 narrative; Representative studies listed are illustrative, not exhaustive

reflection, but has not yet been field-tested with students. Joshi (2025) added a conceptual policy synthesis connecting United States federal initiatives, state implementation, and international benchmarks into an elementary, middle, and high school architecture.

Okada et al. (2025) linked the UNESCO AI competencies framework with the CARE-KNOW-DO pedagogical model and field-tested 70 open educational resources with 330 secondary students in the United Kingdom, Greece, and Brazil. Yu (2025) proposed a competency perspective framework integrating knowledge, skills, dispositions, and context across K-12. Busuttil (2025) developed a constructionist pedagogical framework grounded in Papert's learning-by-making tradition, with age-differentiated recommendations for primary and secondary learners.

This cluster offers system-level organization and competency mapping across grade bands. Its main weakness is the lack of longitudinal validation. Few frameworks track student progression from elementary through secondary school; most rely on cross-sectional data, expert consensus, conceptual synthesis, or short-term implementation evidence.

Primary/elementary-focused frameworks

Six studies focused on elementary or primary learners, making this cluster much smaller than the secondary-focused group and pointing to a thinner evidence base for children under 12. For instance, Yim (2024) proposed an intelligence-based AI literacy framework for primary students that links AI thinking with digital literacy, computational thinking, data literacy, and AI ethics through a post-humanist orientation. Yim and Su (2025) reviewed 25 empirical studies and found that primary AI literacy education relied mainly on constructionist and constructivist pedagogies, with limited validated assessment tools for young learners

Upper-elementary evidence came from three empirical studies. Cheng et al. (2025) adapted the ABCE model for

Taiwanese fifth graders through a 19-week quasi-experimental STEAM-PBL intervention and tested gender measurement invariance. Dai (2024) used a dual-contrast pedagogical approach with 147 sixth graders, combining cognitive conflict with human-AI analogy. Yue et al. (2025) proposed the students as AI literate designers framework and applied design-based learning with 50 grade 5 students. Lin and Dai (2025) added a constructionist approach grounded in epistemic insight theory for upper-primary students aged 10–12.

Across this cluster, the evidence suggests that age-appropriate AI literacy instruction can work in elementary classrooms, especially through hands-on, constructionist, and design-based pedagogies. Assessment and progression evidence remain limited: most studies relied on pre/post knowledge gains or qualitative artefacts, and none tracked development across multiple primary grades.

Secondary-focused competency and assessment frameworks

As for the secondary-focused competency and assessment frameworks, we have sixteen studies focused on middle and high school contexts which makes it the largest and most methodologically varied cluster. It provides the corpus's strongest psychometric evidence, while also exposing a recurring limitation: many instruments assess self-reported awareness, confidence, or perceived competence, not demonstrated AI understanding.

The ABCE framework anchored several studies. Ng et al. (2024c) developed and validated the AI literacy questionnaire (AILQ) with 363 Hong Kong junior secondary students. Sun and Chan (2025) translated and validated the AILQ with 1,669 senior secondary students in Shanghai, then used latent profile analysis to identify four AI literacy profiles. Ng et al. (2024a, 2024b) extended this line through Secondary 1 classroom interventions in Hong Kong.

Scale-development work added further measurement depth. Ahmad et al. (2025) validated a 32-item AI competence scale with 608 high school students in Qatar. Luo et al. (2025) proposed an AI information literacy

framework and validated it with 1,132 Chinese high school students. Şahin et al. (2025) developed the 27-item generative artificial intelligence awareness scale (GAIAS) with 888 Turkish secondary students, using separate exploratory and confirmatory factor-analysis samples. It is one of the first instruments in the corpus focused on adolescent GenAI awareness.

Other studies widened the cluster through subject-specific, ethical, and implementation work. Yang et al. (2024) piloted an AI-STEM project design framework with 64 middle school students in China. Choi et al. (2024) built a Korean middle-school AI ethics program around the Seoul PACT framework. Park and Kwon (2024) developed an AI program aligned with Korea's free semester policy. Tang et al. (2025) adapted a critical questioning framework to GenAI-mediated learning with grade 10 students in Australia, and Adisa and Adefisayo (2025) examined middle school GenAI adoption through UTAUT theory.

This body of research also included curriculum, review, and participatory auditing work. Barber et al. (2024), for example, contributed the Irish transition year AI in my life program. Bosarge (2024) reviewed high school AI literacy studies through the AI4K12 five big ideas lens while Morales-Navarro et al. (2025) developed a five-step algorithm auditing process for high school students, and Levy-Nadav et al. (2025) examined secondary AI literacy through digital literacy, AI literacy, and digital competence frameworks. Overall, this cluster's main contribution is measurement. Its main weakness is the continued reliance on self-report: few studies connect measured competence to observable classroom performance or transferable AI reasoning.

Pedagogical, implementation, and design-based frameworks

Eleven studies translated AI literacy frameworks into classroom modules, prototypes, design activities, or implementation studies. This cluster foregrounds what students do with AI: design, audit, build, classify, reflect, and revise.

RAICA was the most fully documented implementation framework. Stump et al. (2025) described its six pedagogical phases: play, brainstorm, define, pause/plan, prototype, and try out. Masla et al. (2025) reported the picture this formative assessment package for middle school learners, and Bosch et al. (2025) examined RAICA implementation fidelity across beta-test sites in the United States, Chile, Malawi, and the United Arab Emirates. Across the three papers, RAICA connects curriculum design, formative assessment, and multi-site implementation evidence.

A second strand embedded AI literacy in subject areas. Smyslova et al. (2025) taught foundational GenAI concepts through a Markov chains module in grade 9 English language arts. Zha et al. (2025) integrated four AI literacy lessons into high school biology. Dong et al. (2025) used the StoryQ platform across five machine-learning lessons and tracked conceptual change in grades 8, 9, and 11. These studies show that AI literacy can be built into existing subjects.

Other studies focused on curriculum format, interaction, and access. Gelder et al. (2025), for instance, co-designed a middle school AI elective over three years, grounding it in the AI4K12 five big ideas and revised Bloom's taxonomy. On their part, Maljur et al. (2025) developed the AI exploratorium, a gamified exhibition where students trained image-recognition models and examined algorithmic bias while Williams et al. (2024) designed Doodlebot, a constructionist educational robot for primary and middle school learners.

Two studies centered inclusive design. Ojeda-Ramirez et al. (2024) proposed translanguaging-based strategies for multilingual elementary and middle school learners. Mannila et al. (2025) co-designed three human-centered AI lesson plans for grade 3-grade 6 in Finland through a four-stage process with 77 stakeholders. Collectively, this cluster provides the corpus's clearest classroom-ready material. Its limits are equally clear: many studies were pilots or beta tests with short durations, context-specific samples, and limited evidence of long-term transfer. Scalability remains unresolved.

RQ2. What Competencies, Critical Constructs, and Progression Models are Defined?

Competency structures

Starting with competency structures, across the 50 studies, competency definitions converged around cognitive, affective, behavioral, ethical, and metacognitive dimensions, although labels varied. Cognitive competence was the most visible. Ng et al. (2024c) operationalized it in the AILQ through lower-level know/understand items and higher-level apply/evaluate/create items. Later ABCE-aligned studies adapted this structure in validation and intervention work (Cheng et al., 2025; Ng et al., 2024a, 2024b; Sun & Chan, 2025). Zhou et al. (2025) used a different structure, analyzing K-12 AI literacy across technology, collaboration, impact, ethics, and self-reflection. Technology was the most represented area in their reviewed corpus; self-reflection received no coverage.

Affective competence appeared through constructs such as motivation, self-efficacy, attitudes, confidence, and interest. Sun and Chan (2025) found that students in the highest AI literacy profile scored strongly across affective, behavioral, cognitive, and ethical dimensions, with the reverse pattern in the lowest profile. Ahmad et al. (2025) placed affective and human-centered constructs within their human-centered mindset dimension. Career interest proved less stable: Ng et al. (2024c) removed it after expert and teacher review suggested that junior secondary students, with a mean age of 13.1, were too young for reliable career-related judgments.

Behavioral competence received less direct empirical attention. It appeared as collaboration, behavioral commitment, AI interaction, or applied AI use. Park and Kwon (2024), for instance, reported descriptive improvement in students' interaction with AI, but the change was not statistically significant. Chiu et al. (2025)

offered more detailed behavioral dimensions, including multimedia creation with AI and Decision Making with AI. Implementation studies made behavioral competence visible through tasks: Smyslova et al. (2025) used Markov chain modelling; Dong et al. (2025) used classifier training and evaluation; and Gelder et al. (2025) assessed applied work on autonomous robots, perception, route finding, and language-based AI.

As for metacognitive and epistemic competence, this appeared in studies that asked students to reason about AI's nature, limits, and relation to human cognition. Tang et al. (2025) addressed this through a critical questioning framework for GenAI-mediated learning while Şahin et al. (2025) measured related awareness through the GAIAS, and Adisa and Adefisayo (2025) documented middle school students' emerging ability to evaluate AI outputs.

Critical constructs: Ethics, bias, and fairness

Ethics was the most common critical construct, though depth varied. Ma et al. (2025) provided a detailed synthesis, extracting 188 pedagogical design instances from 68 K-12 AI ethics education studies and mapping them onto seven Responsible AI pillars. The pattern suggests stronger classroom attention to fairness, transparency, privacy, and social well-being, with far less coverage of technical robustness and human agency.

Several studies treated bias as a core learning target. Morales-Navarro et al. (2025), for instance, engaged 14 high school students in a participatory audit of TikTok's GenAI, focusing on racial, gender, and age bias. Masla et al. (2025) embedded data bias in an image-classification curriculum, and Bosch et al. (2025) traced how responsible AI and bias-related components were enacted across implementation sites. Choi et al. (2024) built an eight-session Korean middle-school AI ethics program around the Seoul PACT principles, and Maljur et al. (2025) addressed algorithmic discrimination through the AI exploratorium.

Primary and multilingual contexts extended this work. Lin and Dai (2025) used Scratch, Teachable Machine, and Poe with 34 fifth-grade students to build epistemic insight into AI ethics and transparency. Ojeda-Ramirez et al. (2024) added an equity lens through translanguaging-based AI literacy instruction for multilingual elementary and middle school learners, linking bias to algorithmic injustice and minoritized students' access.

Progression models

Developmental progression across grade bands was rare. Several studies proposed sequences, but few provided evidence for the sequence itself. Chung et al. (2025) offered a nuanced three-tier K-12 progression linked to assessment evidence: elementary students recognize AI in everyday life, middle school students study AI's dependence on data, and high school students engage with neural networks, foundation models, and societal implications. Xia (2025), drawing on Cao et al. (2025) as cited in Xia (2025), reported a similar Chinese primary-secondary progression from

perceptual experience to core knowledge and project-based work with advanced applications. Joshi (2025) proposed a US-aligned elementary, middle, and high school progression. Tadimalla and Maher (2024, 2025) offered a Bloom's-based sequence from foundational knowledge to application, evaluation, and creation, but without empirical validation of grade-band transitions.

Most studies offered within-course progression, moving from foundational lessons to application and creation. Cross-grade developmental sequences were largely absent. Okada et al. (2025) scaffolded lower, middle, and upper secondary learning through CARE-KNOW-DO, aligned with UNESCO's understand-apply-create levels, though implementation evidence was reported within each level. Ma et al. (2025) proposed a know/understand, use/apply, and evaluate/create progression, with ethical outcomes ranging from recognizing bias origins to identifying bias in applications and building equitable AI systems. Sun and Chan (2025) confirmed measurement invariance across years 10-12 in the Chinese AILQ, supporting consistent measurement across grades but not developmental growth. No study tracked the same students across multiple grade bands over time. Current progression models remain theoretical, cross-sectional, or curriculum-based. They specify plausible sequences, but longitudinal evidence on how K-12 AI literacy develops over time is still absent.

RQ3. What Assessment Strategies are Proposed or Validated?

Assessment approaches across the 50 studies fell into five broad categories: self-report Likert scales, knowledge tests, multidimensional questionnaires, performance-based or formative assessments, and qualitative measures. The corpus leaned strongly toward self-report instruments, which limits what can be inferred about students' demonstrated AI understanding. **Table 3** summarizes the main assessment approaches and the forms of evidence they generated.

Among validated instruments, self-report scales dominated. Ng et al. (2024c) validated the 32-item AILQ, and Sun and Chan (2025) cross-validated its Chinese version. Ahmad et al. (2025) developed a 32-item AI competence scale; Chiu et al. (2025) validated the 32-item SAICS scale; Luo et al. (2025) produced a 24-item AI information literacy scale; and Şahin et al. (2025) developed the 27-item GAIAS. These instruments report strong psychometric properties, but they mainly assess perceived competence, awareness, or self-efficacy, not demonstrated AI performance. Chiu et al. (2025) made this limitation explicit for SAICS.

Knowledge-based assessments were less common but methodologically important. Chung et al. (2025) developed a 24-item criterion-referenced diagnostic tool validated through item response theory. Dai (2024) used a 32-item AI knowledge test with Scratch project scoring on a 27-point scale. Ng et al. (2024b) combined a 20-item knowledge test with a creativity measure. Smyslova et al. (2025) and Zha et al. (2025) both used pre/post knowledge assessments

Table 3. Assessment approaches and evidence types identified across the corpus

Assessment approach	Representative papers	Evidence type
Self-report Likert scales	Ng et al. (2024c), Sun and Chan (2025), Chiu et al. (2025), Şahin et al. (2025)	Psychometric validation, including CFA/EFA, Cronbach's alpha, KMO/Bartlett tests, and measurement invariance
Knowledge-based tests	Chung et al. (2025), Dai (2024), Ng et al. (2024b)	Multiple-choice tests, diagnostic tools, criterion-referenced items, IRT, reliability checks
Multidimensional questionnaires	Yang et al. (2024), Ahmad et al. (2025), Luo et al. (2025)	Measures of knowledge, practice, attitudes, ethics, information literacy, and interdisciplinary dimensions; reliability and factor-structure evidence
Performance-based/formative assessments	Masla et al. (2025), Stump et al. (2025), Bosch et al. (2025), Yue et al. (2025), Morales-Navarro et al. (2025)	Exit tickets, rubrics, design artefacts, classroom observations, algorithm audits, implementation fidelity data
Qualitative measures	Barber et al. (2024), Levy-Nadav et al. (2025), Lin and Dai (2025), Yue et al. (2025)	Interviews, focus groups, observations, video/screen recordings, artifact analysis, thematic analysis

Note. Assessment categories are analytic groupings used in this review and are based on dominant assessment function; Several studies combined more than one assessment approach

within subject-area instruction and reported significant gains. Yang et al. (2024) added a multidimensional 41-item questionnaire covering technical practice, interdisciplinary thinking, and ethical attitudes.

Performance-based and formative assessment appeared mainly in design-based studies. Masla et al. (2025) developed exit tickets, a teacher-facing analytic rubric, and a student-facing single-point rubric for RAICA. Iterative testing with teachers and students across four continents found that most items had instructional utility, with language clarity driving many revisions. Bosch et al. (2025) found that RAICA assessments were frequently modified during implementation, raising questions about portability. Likewise, Stump et al. (2025) used ICAP interval observations to classify student engagement while Gelder et al. (2025) contributed artefact-based assessment aligned with revised Bloom's taxonomy.

Qualitative methods usually served as complementary evidence: interviews (Barber et al., 2024; Levy-Nadav et al., 2025), focus groups (Yue et al., 2025), video or screen recording analysis (Lin & Dai, 2025), and artefact-based evaluation (Morales-Navarro et al., 2025). Few studies relied on qualitative assessment alone.

Two reviews identified broader assessment gaps. Zhou et al. (2025) noted the absence of a validated comprehensive K-12 AI literacy test covering all key learning topics. Su et al. (2025) reviewed assessment tools across 36 studies and found that 63.8% reported reliability evidence, while fewer than 20% reported validity evidence. Overall, existing instruments assess parts of AI literacy: AILQ captures affective, behavioral, cognitive, and ethical dimensions through self-report; Chung et al.'s (2025) diagnostic tool assesses cognitive understanding; and RAICA rubrics and exit tickets assess responsible design, ethical reasoning, and applied AI understanding. No single instrument spans the full construct space with validity evidence across grade bands.

RQ4. What Evidence Exists for Feasibility and Adoption?

As for feasibility evidence, it ranged from single-classroom pilots to multi-site implementations involving large numbers of teachers and students. The strongest evidence concerns short-term classroom feasibility;

sustained adoption at scale remains less developed, especially around teacher preparation, fidelity, and institutional support.

Several studies reported broad reach. Barber et al. (2024), for example, reported that the AI in my life program reached 92 schools and 3,052 students in Ireland through an ambassador model led by AI researchers and data science students. Okada et al. (2025) documented EU CONNECT implementation with more than 1,500 teachers and 50,000 students across multiple countries, though detailed outcome data were reported for 330 students. The RAICA studies offered a multi-site model linking curriculum design, formative assessment, and fidelity evidence (Bosch et al., 2025; Masla et al., 2025; Stump et al., 2025). Bosch et al. (2025) reported a beta test with 12 educators and approximately 282 students across nine sites in four countries. Activities were implemented "as prescribed" 40.7% of the time and assessments 37.5% of the time, showing substantial teacher adaptation.

Mid-scale studies combined intervention and measurement evidence. Cheng et al. (2025) tested a 19-week intervention with matched experimental and comparison groups. Ng et al. (2024b) implemented a 12-lesson maker-based curriculum and reported significant gains across all four cognitive domains, with Cohen's *d* values from 0.50 to 0.94. Luo et al. (2025) validated an AI information literacy instrument across 11 schools in Shanghai and Anhui Province, while Chiu et al. (2025) validated SAICS across six Hong Kong secondary schools with 448 students.

Smaller pilots provided proof-of-concept evidence. Yang et al. (2024) ran a 16-session after-school program with 64 middle school students and reported significant gains in technical practice. Park and Kwon (2024) implemented an eight-week program with 23 students within Korea's free semester structure. Dai (2024) showed that ten weeks of instruction could produce gains within China's after school 330 project.

Subject integration offered another route to feasibility. Smyslova et al. (2025) delivered a Markov chains module in grade 9 English language arts. Zha et al. (2025) embedded four AI literacy lessons in advanced biology while Dong et al. (2025) used StoryQ across five lessons with teacher

professional development. These studies show that AI literacy can be built into existing subjects, including language arts and biology.

Many studies used low-barrier tools, including Google teachable machine, Scratch with AI extensions, MIT app inventor, and micro:bit. Williams et al. (2024), for instance, addressed resource constraints through Doodlebot, a low-cost educational robot. Ojeda-Ramirez et al. (2024) proposed strategies for multilingual learners using storytelling, role-play, and programming, showing how AI literacy can support linguistic access while preserving conceptual depth.

Teacher preparation remained a recurring constraint. Gelder et al. (2025) documented a three-year researcher-educator co-design process, and Mannila et al. (2025) used a four-stage participatory process with teachers, IT educators, and teacher trainers. On their part, Levy-Nadav et al. (2025) found that teachers' emphasis on some competencies declined over six to eight months. Chiu et al. (2024) noted that many K-12 teachers do not study AI during initial teacher education, and Joshi (2025) reported a 45-point gap between teacher awareness of AI and readiness to teach it. Across these studies, accessible materials helped, but teacher confidence and institutional support remained central barriers beyond pilot settings.

RQ5. What Limitations and Gaps Characterize the Current Evidence Base?

Across the 50 studies, limitations clustered around methodology, construct definition, equity, and assessment alignment. These gaps recurred often enough to suggest structural weaknesses in the current evidence base.

Methodological imitations

Sample constraints were common. Many empirical studies used convenience samples from single schools, programs, or small pilots, with samples ranging from seven participants in Maljur et al. (2025) to 34 students in Lin and Dai (2025). Even large samples did not always solve this problem; Sun and Chan (2025), for example, included 1,669 students from one Shanghai school.

Weak comparison conditions also limited causal claims. Single-group pre/post studies (Choi et al., 2024; Dong et al., 2025; Park & Kwon, 2024; Smyslova et al., 2025; Yang et al., 2024; Zha et al., 2025) cannot isolate intervention effects from maturation, testing effects, or concurrent instruction. Cheng et al. (2025) and Dai (2024) used stronger comparison designs, though causal precision remained limited. No study tracked student outcomes longitudinally across grade bands. Most evidence came from short pre/post windows, cross-sectional profiles, or retrospective interviews. Levy-Nadav et al. (2025) used a two-time-point design over six to eight months, but examined teacher practices, not student outcomes. The corpus therefore offers little evidence on persistence of AI literacy gains or on grade-band progression over time.

Conceptual and definitional gaps

The corpus lacks construct-level consensus. ABCE, the five-component model of technology, impact, ethics, collaboration, and self-reflection, and four-dimensional models such as those used by Ahmad et al. (2025) and Luo et al. (2025) organize AI literacy differently. Each model has internal logic, but labels and dimensions do not map cleanly across studies.

Self-reflection was especially weak. Zhou et al. (2025) found no coverage of self-reflection in their reviewed studies, and the present corpus shows limited operationalization. Few studies specified how students should examine their AI use, monitor dependence on AI tools, or regulate engagement with AI-generated output.

GenAI also remained unevenly represented. Many instruments, including the AILQ, SAICS, and Chung et al.'s (2025) diagnostic tool, were developed around machine learning, computer vision, and natural language processing. Competencies now central to student AI use, such as prompt design, output evaluation, hallucination detection, and appropriate attribution, were less developed. Şahin et al. (2025) developed the only GenAI-specific awareness instrument. Tadimalla and Maher (2025), Adisa and Adefisayo (2025), and Tang et al. (2025) also addressed GenAI competencies, though the evidence remains preliminary or limited to single-classroom contexts.

Equity and representation gaps

Geographic concentration limits cultural transferability. The United States, Hong Kong, and mainland China accounted for 70% of included studies, many from well-resourced, STEM-oriented, or technology-rich contexts. Far less is known about under-resourced schools, rural communities, limited-connectivity settings, or systems with different pedagogical traditions. Bosch et al. (2025) extended beta testing to varied sites, including a nonprofit community development center in a refugee camp context in southeastern Africa (e.g., Malawi), but this was a pilot deployment, not a systematic adaptation study.

Participant-level equity received limited attention. Gender was the most common demographic variable: Cheng et al. (2025) tested gender measurement invariance, and Sun and Chan (2025) examined gender as a profile predictor. Socioeconomic status, disability, linguistic diversity, and race/ethnicity were rarely treated as analytic categories. Ojeda-Ramirez et al. (2024) provided the most sustained engagement with linguistic diversity, and Stump et al. (2025) designed RAICA with accessibility for learners with language-based disabilities and vision impairments.

Assessment-curriculum alignment gap

Assessment and curriculum often developed separately. Scale-development studies (Ahmad et al., 2025; Chiu et al., 2025; Şahin et al., 2025) reported strong cross-sectional psychometric evidence but did not show that their instruments detect instructional effects or growth. Curriculum-focused papers (Chiu et al., 2024; Tadimalla &

Maher, 2024, 2025) proposed learning experiences without validated assessments aligned to their competency structures. Gelder et al. (2025) offered richer assessment examples without full psychometric validation. Su et al. (2025) confirmed this pattern at the review level, finding that fewer than 20% of studies reported validity evidence for assessment tools. Chung et al. (2025) and Masla et al. (2025) were among the few studies that developed framework and assessment together. This integrated design remains uncommon.

DISCUSSION

The 50 papers reviewed here show rapid framework-level growth in K-12 AI literacy, with 35 published in 2025 and 15 in 2024. This pace signals strong interest, but limited consolidation. Researchers have developed multi-component frameworks around cognitive, affective, behavioral, ethical, and social dimensions, yet the evidence base remains fragmented. Frameworks develop in parallel, assessments are often separated from curriculum, and progression claims still rely mainly on theory or cross-sectional evidence.

Across the five RQs, the central tension is clear: individual studies are becoming more sophisticated, but the field lacks shared structures for cumulative work. Chung et al. (2025) and Masla et al. (2025) show that rigorous, grade-banded, assessment-aligned AI literacy work is possible. Both scored 5/5 in our checklist because they addressed grade bands, competencies, critical constructs, assessment, and classroom-ready exemplars together. Most studies addressed only some of these elements, producing frameworks with conceptual value but limited curriculum readiness.

Fragmentation and the Need for Convergence

The field's main structural problem is the proliferation of dimensional models that do not map cleanly onto one another. The ABCE framework began with Ng et al. (2024c) in Hong Kong and was later adapted or validated in Shanghai (Sun & Chan, 2025) and Taiwan (Cheng et al., 2025). Chiu et al. (2025) pursued a related line through an eight-dimension AI competency framework and the 32-item SAICS. Ahmad et al. (2025) proposed four UNESCO-adapted dimensions, and Luo et al. (2025) proposed five AI information literacy dimensions. These models have internal coherence, but their labels and constructs do not align. Cumulative progress depends on clearer links among cognitive, affective, behavioral, ethical, and human-centered dimensions.

The ABCE family is the closest current approximation to a shared measurement tradition, with adaptation or validation across three educational systems. Ma et al. (2025) and Zhou et al. (2025) attempted higher-order synthesis but generated new organising structures without a crosswalk among existing models. A field-level consensus process, similar in purpose to AI4K12's content standards work, could

stabilize competency measurement and support comparison across studies.

The GenAI Gap

Many frameworks in this corpus draw on pre-generative reference technologies: machine learning classifiers, computer vision systems, and natural language processing tools that students encounter as observers, trainers, or evaluators. Competencies now common in schoolwork, including prompt design, output evaluation, hallucination detection, attribution, and recognition of homogenization in AI-mediated writing, remain peripheral. Şahin et al. (2025) developed the only generative-AI-specific instrument in the corpus, and the dominant instruments, AILQ, SAICS, and the Chung et al. (2025) diagnostic tool, do not assess GenAI literacy at all.

To read this as an oversight by individual authors would miss what produces it. Framework development is slow work. Design, expert consensus, piloting, psychometric validation, and publication together take two to three years, which means an instrument published in 2025 encodes a conception of AI formed in 2022 or 2023. The gap identified here is structural, and it will recur. Frameworks now entering validation will meet classrooms already working with agentic and multimodal systems that those frameworks do not describe.

This points to a design principle for the next generation of frameworks. Competencies that survive technological change, such as evaluating outputs against external evidence, reasoning about data provenance, and judging when automated assistance is appropriate, should form the stable core. Competencies tied to the behavior of one system generation belong at a separate level that can be revised without rebuilding the whole structure. None of the frameworks reviewed here makes that separation explicit.

Assessment Without Alignment

Assessment and curriculum often developed separately. Scale-development papers (Ahmad et al., 2025; Chiu et al., 2025; Ng et al., 2024c; Şahin et al., 2025) reported cross-sectional psychometric evidence, showing that their constructs were measurable and internally consistent. They did not show whether scores respond to instruction, detect growth, or distinguish students with AI literacy instruction from those without it. Su et al. (2025) found the same issue at review level: fewer than 20% of studies reported validity evidence for assessment tools.

Curriculum-development papers followed the opposite pattern. Stump et al. (2025), Smyslova et al. (2025), Zha et al. (2025), and Dong et al. (2025) assessed learning through lesson-linked tools such as pre-/post-test, qualitative coding, artefacts, exit tickets, or classroom tasks. These measures show learning within a module but limit comparison across studies. Chung et al. (2025) developed framework and assessment together, and Masla et al. (2025) aligned classroom assessments directly to an existing competency framework, offering stronger models for future work.

Geographic Concentration and Equity

The United States (32%), Hong Kong (26%), and mainland China (12%) accounted for 70% of included studies, raising questions about cultural transferability. Many studies came from well-resourced, STEM-oriented, or technology-rich settings. Applicability to under-resourced schools, rural communities, and educational cultures with different pedagogical traditions remains largely untested. Bosch et al. (2025) extended RAICA to nine pilot sites across four countries, including a nonprofit community development center in a refugee-camp context in Malawi. Okada et al. (2025) worked across the United Kingdom, Greece, and Brazil. Both broadened implementation contexts, but neither conducted systematic contextual adaptation.

Within-sample equity received less attention. Several instruments tested gender invariance or gender differences (Cheng et al., 2025; Chiu et al., 2025; Sun & Chan, 2025), but socioeconomic status, disability, linguistic diversity, and race/ethnicity were rarely analyzed. Ojeda-Ramirez et al. (2024) engaged linguistic diversity, and Bosch et al. (2025) treated accessibility as a design concern through universal design for learning. Future frameworks need more diverse populations and contexts from the outset.

Implications For Practice

The findings point to three priorities. First, AI literacy can be taught across grade bands with low-barrier tools such as Google teachable machine, Scratch, and MIT app inventor, supported by design-based or constructionist pedagogy. Subject-integration studies (Dong et al., 2025; Smyslova et al., 2025; Zha et al., 2025) show that AI literacy can be built into existing courses, including English language arts and biology.

Second, teacher preparation remains a major adoption barrier. Co-design models (Gelder et al., 2025; Mannila et al., 2025) suggest that sustained researcher-teacher collaboration can build ownership, but this requires institutional support that many school systems have not yet provided.

Third, assessment should be developed alongside curriculum. The assessment-curriculum gap identified in **RQ3** and **RQ5** will continue to limit evidence accumulation until validated constructs, classroom tasks, and instructional sequences are designed together.

Limitations of This Review

This review has limitations. The search was restricted to English-language sources published from 2024 to 2025, which may have excluded earlier or non-English work. The five-criterion checklist was grounded in AI literacy literature, but it reflects our operational definition of a framework-level contribution; relevant papers organized differently may have been excluded. The search strategy may also have missed local policy documents, curriculum materials, or regional frameworks outside indexed venues.

Finally, papers published after the 15 January 2026 search date may address some gaps identified here.

CONCLUSION

This scoping review charted K-12 AI literacy frameworks across 50 studies published in 2024 and 2025. The field has produced substantial framework-level activity in a short period, with research groups developing competency structures, assessment instruments, and curricular models in parallel. The strongest studies show that rigorous, grade-banded, assessment-aligned AI literacy work is possible.

The review also identified structural weaknesses. Frameworks proliferate without shared terminology. Assessment instruments are often disconnected from curricular content. Progression models lack longitudinal evidence. GenAI remains peripheral to many frameworks despite its centrality to students' everyday encounters with AI. The evidence base is geographically concentrated, limiting transferability across educational contexts.

These gaps reflect a field in its first years of structured inquiry. The field now needs convergence: longitudinal data, cross-context comparisons, curriculum-aligned assessments, and shared vocabulary that allows this work to accumulate.

Author contributions: **MK & JW:** conceptualization, investigation, methodology, validation, writing – review & editing; **SS:** investigation, methodology. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: Ethical approval was not required because this study reviewed published research and publicly available documents and did not involve direct recruitment of participants or collection of primary participant data.

AI statement: AI tools were used for limited editorial support in the preparation of this manuscript, including language clarity and stylistic consistency. No GenAI tools were used in data analysis, interpretation, or the development of substantive arguments. The authors retain full responsibility for all content and claims.

Declaration of interest: The authors declare no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

- Adisa, I. O., & Adefisayo, A. O. (2025). Middle school students' perspectives on adopting generative AI in K-12 education. *The Journal of Educational Research*, 118(6), 724–735. <https://doi.org/10.1080/00220671.2025.2510393>

- Ahmad, Z., Sultana, A., Abdul Latheef, N., Siby, N., Sellami, A., & Abbasi, S. A. (2025). Measuring students' AI competence: Development and validation of a multidimensional scale integrating educational psychology perspectives. *Acta Psychologica*, 259, Article 105446. <https://doi.org/10.1016/j.actpsy.2025.105446>
- Aromataris, E., & Munn, Z. (Eds.). (2020). *JBIMES-20-01*. Joanna Briggs Institute. <https://doi.org/10.46658/JBIMES-20-01>
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052-1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Barber, G. L., Grehan, L., Dunne, J., Davis, B., & Qadeer, H. (2024). Towards an educational framework for integrating AI education into second-level education in Ireland: Preliminary insights from a national workshop series on AI ethics and privacy (work in progress). In *Proceedings of the 2024 Conference on Human Centred Artificial Intelligence-Education and Practice* (pp. 29-33). ACM. <https://doi.org/10.1145/3701268.3701275>
- Bosarge, E. (2024). Cultivating tomorrow's innovators: Navigating the landscape of high school AI literacy. In *Proceedings of the 2024 ASEE Annual Conference & Exposition. American Society for Engineering Education*. <https://doi.org/10.18260/1-2--47100>
- Bosch, C. A., Gustafson-Quiett, M. C., Abu Hegly, S., Wharton, S., Masla, J., Guterman, L., Macatantan, C., Klopfer, E., Abelson, H., & Breazeal, C. (2025). Advancing research on equitable AI education through a focus on implementation: Insights from a middle school computer vision module beta-test. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28), 29120-29127. <https://doi.org/10.1609/aaai.v39i28.35184>
- Busuttil, L. (2025). Constructing AI literacy: A hands-on approach for school children. *Informatics in Education*, 24(4), 761-779. <https://doi.org/10.15388/infedu.2025.26>
- Cheng, C.-C., Wang, J.-S., Zhai, X., & Yang, Y.-T. C. (2025). AI literacy and gender equity in elementary education: A quasi-experimental study of a STEAM-PBL-AIoT course with questionnaire validation. *International Journal of STEM Education*, 12, Article 50. <https://doi.org/10.1186/s40594-025-00574-y>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, Article 100171. <https://doi.org/10.1016/j.caeo.2024.100171>
- Chiu, T. K. F., Çoban, M., Sanusi, I. T., & Ayanwale, M. A. (2025). Validating student AI competency self-efficacy (SAICS) scale and its framework. *Educational Technology Research and Development*, 73, 2785-2807. <https://doi.org/10.1007/s11423-025-10512-y>
- Choi, J.-I., Yang, E., & Goo, E.-H. (2024). The effects of an ethics education program on artificial intelligence among middle school students: Analysis of perception and attitude changes. *Applied Sciences*, 14(4), Article 1588. <https://doi.org/10.3390/app14041588>
- Chung, K., Kim, S., Jang, Y., Choi, S., & Kim, H. (2025). Developing an AI literacy diagnostic tool for elementary school students. *Education and Information Technologies*, 30, 1013-1044. <https://doi.org/10.1007/s10639-024-13097-w>
- Dai, Y. (2024). Dual-contrast pedagogy for AI literacy in upper elementary schools. *Learning and Instruction*, 91, Article 101899. <https://doi.org/10.1016/j.learninstruc.2024.101899>
- Dong, L., Jiang, S., Xu, F., Duan, S., & Hsieh, C.-J. (2025). From simplification to sophistication: Secondary students' conceptual change in understanding machine learning model decision making. *British Journal of Educational Technology*, 57(2), 423-442. <https://doi.org/10.1111/bjet.70024>
- Gelder, W., Yu, X., Touretzky, D., Gardner-McCune, C., & Uchidiuno, J. (2025). From lecture hall to homeroom: Co-designing an AI elective with middle school CS teachers. *International Journal of Artificial Intelligence in Education*, 35(4), 1749-1792. <https://doi.org/10.1007/s40593-024-00449-3>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Joshi, S. (2025). Enhancing U.S. K-12 competitiveness for the agentic generative AI era: A framework for integrating AI in U.S. K-12 education. *European Journal of Education Studies*, 12(12), 133-189. <https://doi.org/10.46827/ejes.v12i12.6361>
- Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, 45(3), 298-311. <https://doi.org/10.1080/17439884.2020.1754236>
- Levy-Nadav, L., Shamir-Inbal, T., & Blau, I. (2025). Digital competencies for effective GenAI use in secondary schools: A longitudinal exploration of teachers' perspectives and classroom practices. *Journal of Computer Assisted Learning*, 41(5), Article e70123. <https://doi.org/10.1111/jcal.70123>
- Lin, Z., & Dai, Y. (2025). Fostering epistemic insights into AI ethics through a constructionist pedagogy: An interdisciplinary approach to AI literacy. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28), 29171-29177. <https://doi.org/10.1609/aaai.v39i28.35190>
- Lin, Z., Dai, Y., & Ng, O.-L. (2025). Constructionism in K-12 AI literacy education: A systematic review of pedagogical designs, student outcomes, and learning mechanisms. *Journal of Educational Computing Research*, 63(7-8), 1748-1781. <https://doi.org/10.1177/07356331251360442>

- Liu, X., & Zhong, B. (2024). A systematic review on how educators teach AI in K-12 education. *Educational Research Review*, 45, Article 100642. <https://doi.org/10.1016/j.edurev.2024.100642>
- Livingstone, S., Stoilova, M., & Nandagiri, R. (2020). Data and privacy literacy: The role of the school in educating children in a datafied society. In D. Frau-Meigs, S. Kotilainen, M. Pathak-Shelat, M. Hoechsmann, & S. R. Poyntz (Eds.), *The handbook of media education research* (pp. 413-425). Wiley. <https://doi.org/10.1002/9781119166900.ch38>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). ACM. <https://doi.org/10.1145/3313831.3376727>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Luo, M., Kuang, Z., Ng, D. T. K., Zheng, Z., & Yang, Y. (2025). Exploring the influential factors and key dimensions of AI information literacy among high school students: A social-ecological perspective. *Interactive Learning Environments*, 34(6), 3727-3750. <https://doi.org/10.1080/10494820.2025.2575016>
- Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025). Fostering responsible AI literacy: A systematic review of K-12 AI ethics education. *Computers and Education: Artificial Intelligence*, 8, Article 100422. <https://doi.org/10.1016/j.caeai.2025.100422>
- Maljur, T., Novak, J., Laufer, J., & Gnoth, S. (2025). AI exploratorium: A gamified interactive exhibition for developing critical AI literacy. In *Proceedings of the 19th European Conference on Games Based Learning*. Academic Conferences and Publishing International. <https://doi.org/10.34190/ecgbl.19.2.3974>
- Mannila, L., Hägglund, S., & Wingren, M. (2025). Co-designing lesson plans for supporting human-centred AI literacy in K-12 education. In B. K. Smith, & M. Borge (Eds.), *Learning and collaboration technologies* (pp. 102-118). Springer. https://doi.org/10.1007/978-3-031-93564-0_7
- Masla, J., Bosch, C. A., Ravi, P., Guterman, L., Wharton, S., Gustafson-Quiett, M. C., Abu Hegly, S., Macatantan, C., Klopfer, E., Breazeal, C., & Abelson, H. (2025). Supporting AI literacy teaching through the development of assessments for classroom use. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28), 29178-29185. <https://doi.org/10.1609/aaai.v39i28.35191>
- McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology*, 75, 40-46. <https://doi.org/10.1016/j.jclinepi.2016.01.021>
- Morales-Navarro, L., Gan, M. A., Yu, E., Vogelstein, L., Kafai, Y. B., & Metaxa, D. (2025). Learning AI auditing: A case study of teenagers auditing a generative AI model. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), Article 439. <https://doi.org/10.1145/3757620>
- Ng, D. T. K., Chen, X., Leung, J. K. L., & Chu, S. K. W. (2024a). Fostering students' AI literacy development through educational games: AI knowledge, affective and cognitive engagement. *Journal of Computer Assisted Learning*, 40(5), 2049-2064. <https://doi.org/10.1111/jcal.13009>
- 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>
- Ng, D. T. K., Su, J., & Chu, S. K. W. (2024b). Fostering secondary school students' AI literacy through making AI-driven recycling bins. *Education and Information Technologies*, 29, 9715-9746. <https://doi.org/10.1007/s10639-023-12183-9>
- Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024c). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, 55(3), 1082-1104. <https://doi.org/10.1111/bjet.13411>
- Ojeda-Ramirez, S., Ritchie, D., & Warschauer, M. (2024). AI literacy for multilingual learners: Storytelling, role-playing, and programming. *The CATESOL Journal*, 35(1). <https://doi.org/10.5070/B5.35861>
- Okada, A., Sherborne, T., Panselinas, G., & Kolionis, G. (2025). Fostering transversal skills through open schooling supported by the CARE-KNOW-DO pedagogical model and the UNESCO AI competencies framework. *International Journal of Artificial Intelligence in Education*, 35, 1953-1998. <https://doi.org/10.1007/s40593-025-00458-w>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253. <https://doi.org/10.1518/001872097778543886>
- Park, W., & Kwon, H. (2024). Implementing artificial intelligence education for middle school technology education in Republic of Korea. *International Journal of Technology and Design Education*, 34, 109-135. <https://doi.org/10.1007/s10798-023-09812-2>
- Peters, M. D. J., Godfrey, C. M., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Scoping reviews. In E. Aromataris, & Z. Munn (Eds.), *JBIM manual for evidence synthesis*. Joanna Briggs Institute. <https://doi.org/10.46658/JBIMES-20-12>

- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, Article 39. <https://doi.org/10.1186/s13643-020-01542-z>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676-688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Rogers, T., & Carbonaro, M. (2025). From understanding to creating: Bridging AI literacy and AI fluency in K-12 education. *Journal of Teaching and Learning*, 19(4), 20-38. <https://doi.org/10.22329/jtl.v19i4.9420>
- Şahin, R. S., Özbay, Ö., Çınar Özbay, S., & Durmuş Sarıkahya, S. (2025). Development of the generative artificial intelligence awareness scale for secondary school students in Türkiye. *European Journal of Pediatrics*, 184, Article 585. <https://doi.org/10.1007/s00431-025-06435-8>
- Smyslova, D., Jiang, S., Rosé, C. P., Ellis, R., Chao, J., & Li, Q. (2025). Introducing generative AI with Markov chains: Gendered patterns of competence in English language arts classrooms. *The Journal of Educational Research*, 118(6), 701-715. <https://doi.org/10.1080/00220671.2025.2510409>
- Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives*, 14(3), 171-177. <https://doi.org/10.1111/cdep.12381>
- Stump, G. S., Kang, S., Masla, J., Bosch, C. A., Abelson, H., Klopfer, E., & Breazeal, C. (2025). Ethical thinking: Integration and measurement in an AI curriculum for middle-high school students. *Informatics in Education*, 24(4), 815-844. <https://doi.org/10.15388/infedu.2025.29>
- Su, J., Chen, X., Chu, S. K. W., & Hu, X. (2025). A scoping review of empirical research on AI literacy assessments. *Educational Technology Research and Development*, 73, 3105-3130. <https://doi.org/10.1007/s11423-025-10515-9>
- Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, 4, Article 100124. <https://doi.org/10.1016/j.caeai.2023.100124>
- Sun, Z. L., & Chan, Y. Y. (2025). AI literacy development in middle school: Profiles and demographic predictors from an ABCE perspective. *Computers and Education: Artificial Intelligence*, 9, Article 100480. <https://doi.org/10.1016/j.caeai.2025.100480>
- Tadimalla, S. Y., & Maher, M. L. (2024). AI literacy for all: Adjustable interdisciplinary socio-technical curriculum. In *Proceedings of the 2024 IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. <https://doi.org/10.1109/FIE61694.2024.10893159>
- Tadimalla, S. Y., & Maher, M. L. (2025). AI literacy as a core component of AI education. *AI Magazine*, 46(2), Article e70007. <https://doi.org/10.1002/aaai.70007>
- Tan, Q., & Tang, X. (2025). Unveiling AI literacy in K-12 education: A systematic literature review of empirical research. *Interactive Learning Environments*, 33(9), 5347-5363. <https://doi.org/10.1080/10494820.2025.2482586>
- Tang, K.-S., Cooper, G., Rappa, N., & Edwards, J. (2025). Critical questioning with generative AI: Developing AI literacy in secondary education. *Thinking Skills and Creativity*, 59, Article 102043. <https://doi.org/10.1016/j.tsc.2025.102043>
- Touretzky, D. S., Gardner-McCune, C., & Seehorn, D. (2023). Machine learning and the five big ideas in AI. *International Journal of Artificial Intelligence in Education*, 33, 233-266. <https://doi.org/10.1007/s40593-022-00314-1>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467-473. <https://doi.org/10.7326/M18-0850>
- UNICEF. (2021). Policy guidance on AI for children 2.0. UNICEF. <https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf>
- Williams, R., Ali, S., Alcantara, R., Burghleh, T., Alghowinem, S., & Breazeal, C. (2024). Doodlebot: An educational robot for creativity and AI literacy. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 772-780). ACM. <https://doi.org/10.1145/3610977.3634950>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing strands, and future patterns in AI in education. *Learning, Media and Technology*, 45(3), 223-235. <https://doi.org/10.1080/17439884.2020.1798995>
- Xia, X. (2025). What to teach and how to teach: Artificial intelligence education in primary and secondary schools. *Science Insights Education Frontiers*, 28(1), 4613-4626. <https://doi.org/10.15354/sief.25.re523>
- Yang, H., Zhang, D., Guo, W., & He, Y. (2024). Constructing a STEM projects design framework for middle school AI courses: Developing students' AI literacy as project goals. *IEEE Access*, 12, 192101-192112. <https://doi.org/10.1109/ACCESS.2024.3512499>
- Yim, I. H. Y. (2024). A critical review of teaching and learning artificial intelligence (AI) literacy: Developing an intelligence-based AI literacy framework for primary school education. *Computers and Education: Artificial Intelligence*, 7, Article 100319. <https://doi.org/10.1016/j.caeai.2024.100319>

- Yim, I. H. Y., & Su, J. (2025). Artificial intelligence literacy education in primary schools: A review. *International Journal of Technology and Design Education*, 35, 2175-2204. <https://doi.org/10.1007/s10798-025-09979-w>
- Yu, W. (2025). A conceptual framework for AI literacy with a focus on competency. *International Journal of Artificial Intelligence in Education*, 35(5), 2975-2992. <https://doi.org/10.1007/s40593-025-00488-4>
- Yue, M., Jong, M. S.-Y., Dai, Y., & Lau, W. W.-F. (2025). Students as AI literate designers: A pedagogical framework for learning and teaching AI literacy in elementary education. *Journal of Research on Technology in Education*, 58(3), 593-614. <https://doi.org/10.1080/15391523.2025.2449942>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zha, S., Bragdon, M. M., Gong, N., Wang, J., Leavesley, S., Eaton, R., & Bosarge, E. (2025). A case study of integrating AI literacy education in a biology class. *International Journal of Artificial Intelligence in Education*, 35(4), 2453-2477. <https://doi.org/10.1007/s40593-025-00476-8>
- Zhou, X., Li, Y., Chai, C. S., & Chiu, T. K. F. (2025). Defining, enhancing, and assessing artificial intelligence literacy and competency in K-12 education from a systematic review. *Interactive Learning Environments*, 33(10), 5766-5788. <https://doi.org/10.1080/10494820.2025.2487538>