EDUTECH

Jurnal Teknologi Pendidikan

Journal homepage https://ejournal.upi.edu/index.php/edutech

Human-AI Collaboration for Critical Thinking, Creativity, and AI Literacy in Secondary Education: An Integrative Review and Evidence Appraisal of a Five-Phase Learning Framework

Luthfi Malahati Astianti[1*], Budiyono[1], Titi Prihatin[1], Rafika Bayu Kusumandari[1]

[1]Department of Curriculum Development, Faculty of Education and Psychology, Universitas Negeri Semarang, Semarang, Indonesia

*Correspondence: e-mail: luthfimastianti@students.unnes.ac.id

ABSTRACT

This review synthesizes evidence on human-AI collaboration in learning in relation to critical thinking, creativity, and AI literacy, and tests a five-phase human-AI collaborative learning framework formulated earlier by the authors. An integrative literature review design with structured eligibility criteria was used. Studies published from 2019 to 2026 were identified from the first author's thesis reference library, built from topic-based searches in Consensus and Google Scholar, and from supplementary Google Scholar searches and AI-assisted literature recommendations. Thirty-eight studies were included; the 23 empirical studies were appraised with the MMAT, and all studies were synthesized thematically against six research questions. Generative AI strengthened the three competencies when it was integrated through structured, sequenced designs that keep students in control of their learning. Four analytical themes were identified: (a) pedagogical structure determines the direction of AI's benefits; (b) an activity sequence in which "humans initiate, AI extends, and humans decide"; (c) AI literacy links critical thinking and creativity; and (d) teachers and curriculum materials determine implementation. The critical evaluation phase was supported by empirical evidence with moderate confidence. The creative elaboration phase was supported mainly by evidence from higher education and adult samples. Synthesis and reflection rested on conceptual work, and exploration was partially supported and should begin with students' own ideas. Because study identification was not fully reproducible and was carried out by one reviewer, the findings should be read as an evidence-informed basis for design rather than an exhaustive synthesis.

© 2025 Teknologi Pendidikan UPI

ARTICLE INFO

Article History:
Submitted/Received 7 Sept 2026
First Revised 19 Sept 2026
Accepted 1 Oct 2026
First Available online 31 Oct 2026
Publication Date 31 Oct 2026

Keyword:
AI literacy,
Creativity,
Critical thinking,
Generative AI,
Human-AI collaboration,
Integrative review,
Secondary education.

1. INTRODUCTION

Generative artificial intelligence (AI) has become part of secondary school students' everyday learning. Schools therefore need to prepare students who can use AI, evaluate its outputs critically, create with it, and understand how it works and what its ethical implications are. In Indonesia, this demand is reinforced by the Learning Outcomes (*Capaian Pembelajaran*) for the Coding and Artificial Intelligence (*Koding dan Kecerdasan Artifisial*, KKA) subject, which include an element on the Use and Development of Artificial Intelligence (BSKAP, 2025).

The educational literature reflects a shift in how AI is viewed. AI was once seen as a system that directs students and is now increasingly seen as a partner that empowers them to lead their own learning (Ouyang & Jiao, 2021). The concepts of hybrid intelligence and human-AI co-regulation of learning treat AI as an extension of students' cognitive capacity, while control over learning stays with humans (Cukurova, 2025; Järvelä et al., 2023; Molenaar, 2022). Researchers and international organizations have also developed AI literacy frameworks that describe the competencies students need to interact with AI critically and responsibly (Long & Magerko, 2020; Ng et al., 2021; UNESCO, 2024).

However, empirical evidence on human-AI collaboration is still scattered across separate domains. Research on critical thinking, creativity, and AI literacy has largely developed in parallel. Most of it has been conducted in higher education, and it has rarely been linked to classroom-level curriculum design decisions. Existing systematic reviews have focused on a single outcome, such as AI literacy in K-12 (Casal-Otero et al., 2023) or critical thinking (Liu et al., 2025; Sardi et al., 2025). Pérez-Montesdeoca et al. (2026) reviewed the integration of generative AI in secondary education and recommended integrated, sustained, and theory-based pedagogical models, but they did not synthesize the three competencies jointly or specify how such models should be sequenced at the level of learning activities. As a result, teachers and developers of instructional materials still lack a synthesis that connects these three competencies to sequences of learning activities that can be implemented in the classroom.

The present integrative review differs from these earlier reviews in three respects. First, it synthesizes evidence on critical thinking, creativity, and AI literacy within one analytical frame rather than treating each outcome separately. Second, it translates the synthesis into operational design principles at the level of activity sequences, specifying what students do before, during, and after interacting with AI. Third, it grades the strength of the evidence and its relevance to secondary schools for each phase of a classroom framework, so that design decisions can be weighed against the quality and transferability of the underlying evidence.

This review also builds on the first author's ongoing master's thesis research on the development of KKA teaching modules, which has not yet been published. In that research, human-AI collaboration had already been operationalized into five learning phases before the present review was undertaken: exploration, critical evaluation, creative elaboration, synthesis, and reflection. The phases were derived deductively from four theoretical sources: the AI-empowered paradigm and the concept of hybrid intelligence, which keep learners in control of AI use (Molenaar, 2022; Ouyang & Jiao, 2021); Kolb's (1984) experiential learning cycle, which supplied the cyclical movement from concrete exploration to reflection; Vygotsky's (1978) social constructivism, in which AI functions as a scaffold within the zone of proximal development; and cognitive load theory (Sweller, 1988), which motivated separating evaluation, elaboration, and synthesis into distinct steps. The phases were then aligned with the KKA learning outcomes for the element on the Use and Development of Artificial Intelligence (BSKAP,

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

2025) and refined through supervisory review at the thesis proposal stage. A framework developed before a literature review needs to be tested against the available evidence so that it does not rest on selectively chosen references. Accordingly, this review synthesizes research published between 2019 and 2026 on human-AI collaboration in learning to address the following six research questions.

RQ1. How is human-AI collaboration in learning conceptualized and operationalized into learning roles, phases, and activities?
RQ2. How is generative AI-assisted learning related to students' critical thinking, and which design conditions determine the direction of its effects?
RQ3. How is human-AI collaboration related to students' creativity at the process and product levels?
RQ4. How is AI literacy conceptualized, measured, and developed at the secondary school level?
RQ5. Which principles of curriculum design, instructional materials, and teacher competence support the meaningful integration of AI in the classroom?
RQ6. To what extent does this evidence support the previously formulated five-phase framework for human-AI collaborative learning?

This review makes two contributions as a basis for refining secondary school instructional materials. It synthesizes evidence across all three competencies, and it examines the five-phase framework transparently, including the components that are not yet supported by evidence.

2. METHODS

This study used an integrative literature review design (Whittemore & Knafl, 2005), which allows empirical studies, reviews, and conceptual frameworks to be synthesized together, combined with pre-specified eligibility criteria and a structured appraisal and synthesis procedure. The design was chosen because the research questions require conceptual as well as empirical evidence and because study identification drew on the first author's existing thesis literature base rather than on a single reproducible database search. Reporting follows the PRISMA 2020 items that apply to this design (Page et al., 2021); items that could not be met are stated explicitly. A thematic-narrative synthesis was used because the studies' designs were too heterogeneous to pool effect sizes in a meta-analysis. The review protocol was not registered, because PROSPERO accepts only reviews with a health-related outcome and the review was carried out within a time-bound thesis project. The research questions, the PICOC framework, the eligibility criteria, and the appraisal tool were specified in writing in the first author's thesis proposal before study selection, apart from the clarification of criterion E2 described below. The PRISMA 2020 checklist, annotated for this design, is provided as Supplementary Material S1.

2.1. Review Question Framework

The research questions were formulated using the PICOC framework, as presented in Table 1.

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

Table 1. PICOC framework.

Component	Formulation
Population	Secondary school students as the primary focus; university students, teachers, adult samples, and cross-level studies as comparative evidence
Intervention	Learning that involves student interaction or collaboration with AI, particularly generative AI
Comparison	Learning without AI, AI use without pedagogical structure, or no comparison (conceptual and instrument studies)
Outcome	Critical thinking, creativity, AI literacy, instructional design principles, and teacher competence
Context	Formal education; publications from 2019 to 2026

2.2. Identification of Studies

Studies were identified from two sources (Table 2). The first was the reference library that the first author had assembled in Zotero for the literature review of the thesis. This library was built from topic-based searches in Consensus, conducted on 3 July 2026 for six topics (human-AI collaborative learning, critical thinking in the context of AI, creativity in the context of AI, AI literacy, the relationships among these three competencies, and instructional module development), and from Google Scholar searches whose results were exported to Zotero; it contained 370 items when it was imported on 7 August 2026. The second source consisted of supplementary searches, carried out before the included studies were compiled, to locate recent and secondary school studies not covered by the library: keyword searches in Google Scholar on the review's core concepts, which were not logged, and literature recommendations from generative AI tools. Every study, including each one recommended by an AI tool, was checked for existence and publication status on the publisher's website and read in full, so that no study entered the review on the basis of an AI-generated citation alone.

Table 2. Sources used to identify the included studies.

Source	Search approach and topics	Period	Included studies
Thesis reference library (Zotero)	Consensus topic searches (human-AI collaborative learning; critical thinking in the context of AI; creativity in the context of AI; AI literacy; relationships among the three competencies; instructional modules) and exported Google Scholar results	July–August 2026	12
Supplementary searches	Unlogged Google Scholar keyword searches on the core concepts and literature recommendations from generative AI tools, each verified on the publisher's website	Up to 7 October 2026	26
Total			38

All 38 included studies were compiled in a dedicated Zotero collection on 7 October 2026. Because the supplementary searches were not logged, the identification stage

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

cannot be reproduced exactly. To assess this limitation transparently, a post hoc check was run in Publish or Perish (Harzing, 2007) on 21 October 2026 with three Google Scholar keyword queries built from the review's concepts in English and Indonesian. The exact query strings were not retained; the query dates and all exported records are provided in Supplementary Material S2. The check retrieved 238 unique records, none of which overlapped with the included studies. The included corpus therefore reflects a purposive, criteria-based selection from the thesis literature base and supplementary searches rather than the output of a single structured search, and the review does not claim to be exhaustive.

2.3. Eligibility Criteria

The inclusion and exclusion criteria are presented in Table 3.

Table 3. Inclusion and exclusion criteria.

Code	Inclusion criteria	Code	Exclusion criteria
I1	Published 2019–2026	E1	Preprints, undergraduate or graduate theses, or reports that have not undergone peer review (the official frameworks covered by I2 are exempt)
I2	Peer-reviewed journal articles or conference proceedings; official competency frameworks issued by international organizations	E2	Populations not engaged in learning or teaching; studies of general adult populations are excluded unless they directly test a cognitive mechanism of AI use on critical thinking or creativity
I3	Written in English or Indonesian	E3	Addresses AI only from a technical perspective, without a learning dimension
I4	Addresses student interaction with AI and at least one outcome: critical thinking, creativity, AI literacy, or instructional design and teacher competence	E4	Full text inaccessible, or insufficient data
I5	Empirical study, systematic review, or peer-reviewed conceptual framework	E5	Article retracted by the publisher

Two clarifications apply to these criteria. First, official competency frameworks issued by international organizations do not undergo journal peer review, but they are produced through institutional expert consultation and serve as normative references for curriculum design. One such framework (UNESCO, 2024) was included under I2. It was used only to describe how AI literacy is conceptualized (RQ4) and was not counted as empirical evidence for any phase of the framework; E1 therefore applies to preprints, theses, and reports that have neither peer review nor institutional authority. Second, E2 was clarified during full-text assessment. It excluded populations that were not engaged in learning or teaching. Two studies with general adult populations were retained because they directly tested cognitive mechanisms targeted by this review: the effect of

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

AI-generated ideas on creative output (Doshi & Hauser, 2024) and the association between AI use, cognitive offloading, and critical thinking (Gerlich, 2025). In line with the PICOC framework, which treats non-secondary samples as comparative evidence, both studies are flagged in Table 5, were not used on their own to support claims about secondary school students, and were rated as having low relevance to secondary schools in the evidence review (Table 6).

2.4. Study Selection and Quality Appraisal

Candidate studies from both sources were assessed by the first author against the inclusion criteria (I1–I5) and the exclusion criteria (E1–E5) together, on the basis of their full texts. The number of candidate records assessed and excluded at each step was not recorded. Reasons for exclusion that were encountered included preprint status, a population not engaged in learning or teaching, insufficient data, and retraction by the publisher (one meta-analysis was retracted in May 2026 because of inconsistencies in its analysis). For each included study, publication status was verified on the publisher's website or the institutional repository.

The Mixed Methods Appraisal Tool (MMAT), version 2018 (Hong et al., 2018), was applied only to the 23 empirical primary studies. Preliminary ratings were drafted with the assistance of a generative AI tool and were then checked and, where necessary, revised by the first author against the full text of each study, using the five MMAT criteria for its design category, each rated "yes", "no", or "can't tell". The ratings were converted into three categories: high quality (all five criteria met), moderate quality (three or four criteria met), and low quality (two or fewer criteria met). The final criterion-level ratings for each study, with the justification for each criterion, are provided in Supplementary Material S3. The six systematic or exploratory reviews and the nine conceptual frameworks were not given an MMAT rating, because the tool is not designed for these study types. Their methodological transparency (for reviews) and peer-review status and publisher authority (for frameworks) were described narratively. These studies were used as contextual and comparative evidence and could not, on their own, determine the level of support for any phase of the framework. Low-quality empirical studies were kept in the review, but no conclusion rested on them alone.

2.5. Data Extraction and Synthesis

The extracted data included publication details, study type and design, educational level and country, sample size, intervention or construct, outcomes, and key findings. The synthesis used thematic analysis (Thomas & Harden, 2008) in three steps. Findings were coded, the codes were grouped into descriptive themes for each research question, and analytical themes were then developed and translated into implications for instructional design. To address RQ6, each phase of the framework was compared with the analytical themes and the relevant study findings, and was classified with the decision rules in Table 4.

Table 4. Decision rules for classifying the level of evidence support.

Level of support	Decision rule
Supported by empirical evidence	At least two empirical studies of moderate or high MMAT quality, at least one of which has a comparison group, test activities that correspond directly to the phase, with a consistent direction of findings and no contradictory finding from a high-quality study.

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

Level of support	Decision rule
Supported by conceptual evidence	The phase is consistent with at least two peer-reviewed conceptual frameworks or systematic reviews, but no empirical study, or only one, tests activities that correspond directly to the phase.
Partially supported	Evidence meeting the first or second rule supports part of the phase as formulated, but at least one moderate- or high-quality empirical study contradicts another part of its formulation (e.g., its timing or sequence).
Not yet supported	No relevant evidence, or only low-quality empirical evidence, was found for the phase.

In addition to the level of support, confidence in each classification was rated as high, moderate, low, or very low with an approach adapted from GRADE-CERQual (Lewin et al., 2018). The rating considered four components: the methodological limitations of the contributing studies, the coherence of their findings, the adequacy of the data (number of studies and participants), and the relevance of the evidence to secondary school settings. Confidence started at high and was downgraded by one level for each component with serious concerns. Because this review is concerned with secondary schools, evidence drawn mainly from higher education or adult samples was downgraded for relevance. Evidence that did not fit the framework was reported explicitly.

3. RESULTS AND DISCUSSION

3.1. Study Selection

Thirty-eight studies were included (Figure 1): 12 from the thesis reference library and 26 from the supplementary searches. Figure 1 shows the sources and the stages of selection; the numbers of candidate records assessed and excluded were not recorded and are therefore not reported.

DOI: https://doi.org/10.17509/e.v25i3.106791
P-ISSN 0852-1190 E-ISSN 2502-0781

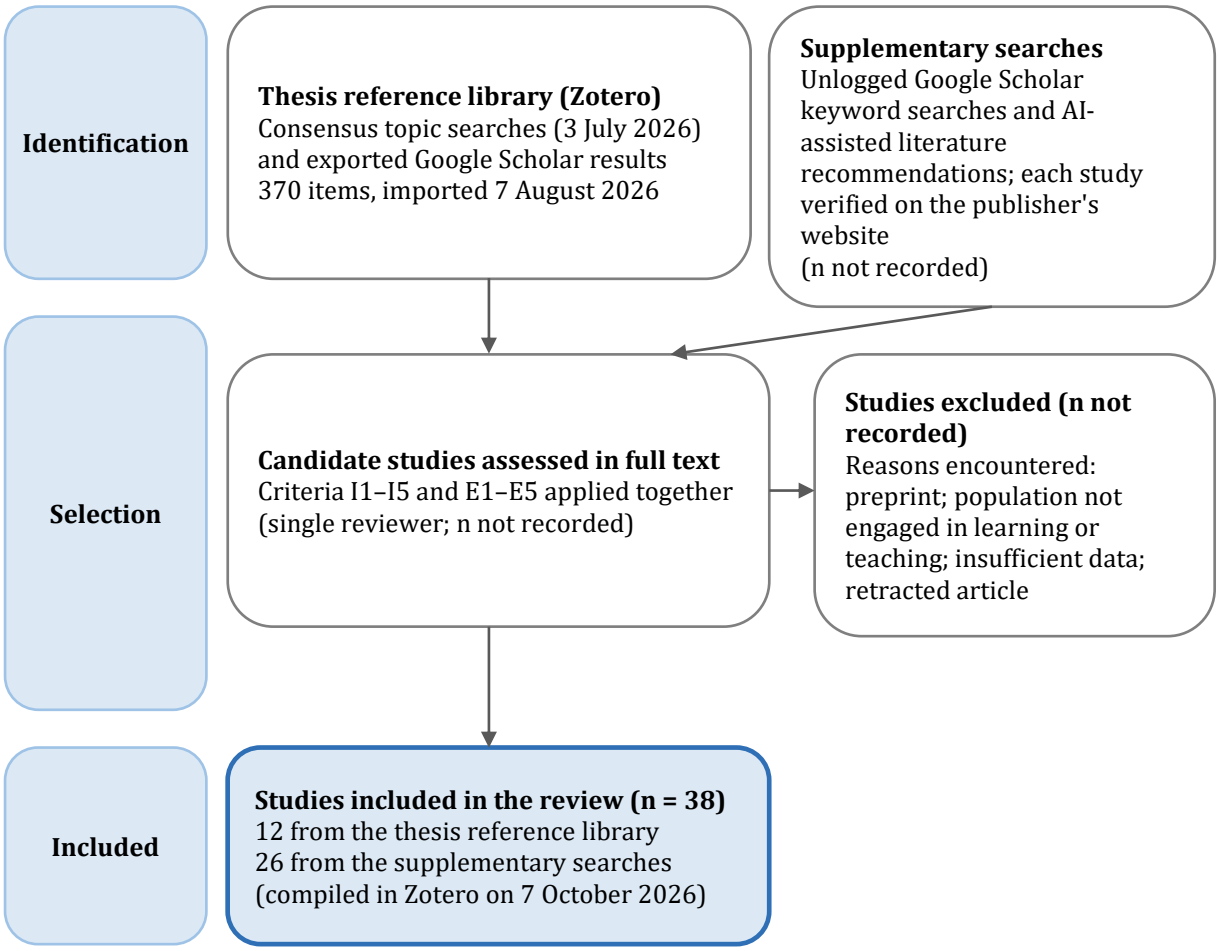


Figure 1. Flow of study identification and selection (adapted from the PRISMA 2020 flow diagram).

3.2. Study Characteristics

Thirty-one of the 38 studies were published between 2023 and 2026 (Figure 2). Most of the evidence on human-AI collaboration in learning has therefore appeared within the past four years, and the body of evidence is still growing.

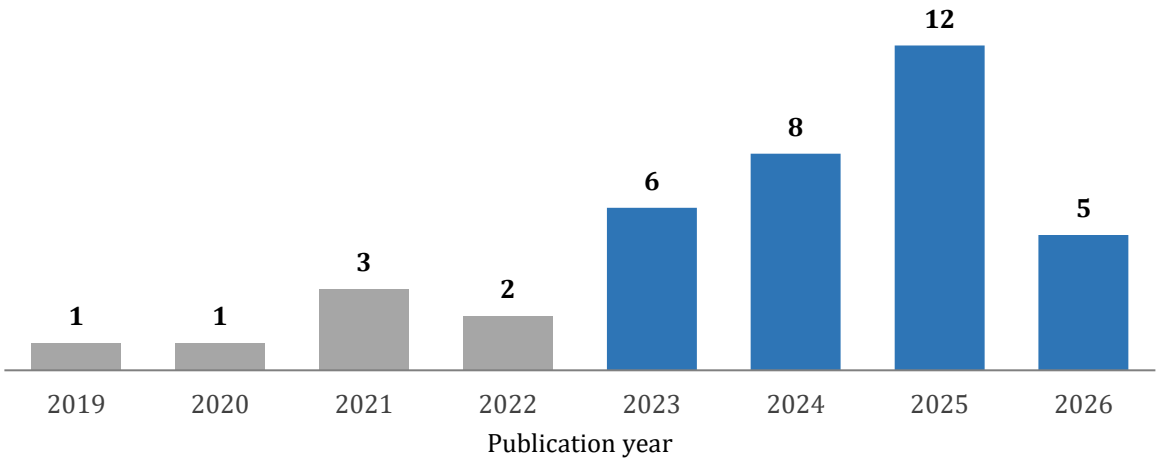


Figure 2. Distribution of studies by publication year (N = 38).

By primary design, the corpus comprised 10 experimental or quasi-experimental studies (including pre-post course and programme evaluations), 8 survey or instrument

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

validation studies, 5 qualitative or mixed-methods studies, 6 systematic or exploratory reviews, and 9 conceptual frameworks. Studies that combined several methods were categorized by the design that produced the findings synthesized in this review. For example, Gerlich (2025) was classified as a survey because its interviews were supplementary, and Kong et al. (2024) and Kong and Yang (2025) were classified as pre-post evaluations because their qualitative data or instrument validation complemented the pre-post comparison. The design category of each study is shown in Table 5. By context, 14 studies were set in K-12 schools, 9 in higher education, 4 focused on teachers, and 11 were cross-level or conceptual. Indonesia was represented by three studies (Nurcahyani & Noviani, 2026; Putri & Ekasriadi, 2026; Sari et al., 2025). Of the 23 empirical studies, 6 were rated high quality, 12 moderate, and 5 low (Supplementary Material S3). Two empirical studies involved general adult populations (Doshi & Hauser, 2024; Gerlich, 2025) and are flagged in Table 5. Table 5 summarizes the design, sample, setting, quality, and key finding of each study; full bibliographic details and DOIs are given in the reference list.

Table 5. Characteristics of the included studies (N = 38).

No	Study	Design category	Design and sample	Level and country	RQ	Quality	Key finding
1	Zawacki-Richter et al. (2019)	Review	Systematic review; 146 articles	HE; international	RQ2, RQ5	n/a	Research on AI in higher education shows little pedagogical reflection and a weak link to educational theory.
2	Long & Magerko (2020)	Concept	Conceptual framework	Cross-level	RQ4	n/a	Defines AI literacy through competencies and design considerations organized under five themes.
3	Ouyang & Jiao (2021)	Concept	Conceptual framework	Cross-level	RQ1	n/a	Distinguishes AI-directed, AI-supported, and AI-empowered paradigms of AI in education.
4	Ng et al. (2021)	Review	Exploratory review	Cross-level	RQ4	n/a	Identifies four aspects of AI literacy: know and understand, use and apply, evaluate and create, and ethics.
5	Seo et al. (2021)	Qual/Mixed	Qualitative (Speed Dating interviews); 12 students, 11 instructors	Online HE	RQ1, RQ5	High	AI can strengthen learner engagement but requires clear allocation of responsibility and control.
6	Kim et al. (2022)	Qual/Mixed	Qualitative interviews; 10 teachers	Teachers; South Korea	RQ1, RQ5	High	Recommends authentic, creative tasks and process-based assessment; collaboration develops from learning about, from, and with AI.
7	Molenaar (2022)	Concept	Conceptual framework	Cross-level	RQ1	n/a	The greatest benefits arise at intermediate levels of automation, where educators keep control.
8	Järvelä et al. (2023)	Concept	Conceptual framework	Cross-level	RQ1	n/a	Proposes hybrid human-AI shared regulation triggered by critical moments in collaboration.
9	Lodge et al. (2023)	Concept	Conceptual framework	HE	RQ1	n/a	Frames learning with generative AI as a network of co-regulation centered on individual agency.

No	Study	Design category	Design and sample	Level and country	RQ	Quality	Key finding
10	Casal-Otero et al. (2023)	Review	Systematic review; 179 documents	K-12	RQ4, RQ5	n/a	Few learning experiences assessed AI literacy after the intervention; recommends modular curricula co-developed with educators.
11	Mishra et al. (2023)	Concept	Conceptual framework	Teachers	RQ5	n/a	Generative AI requires teachers to expand their contextual knowledge within TPACK.
12	Celik (2023)	Survey	Survey and framework validation; 428 teachers	Teachers	RQ5	Moderate	Teachers' technological knowledge becomes meaningful only when combined with pedagogical knowledge.
13	Kasneci et al. (2023)	Concept	Conceptual study	Cross-level	RQ2, RQ5	n/a	A pedagogical framework emphasizing critical thinking and fact-checking is a precondition for integrating LLMs.
14	Essel et al. (2024)	Exp	Randomized experiment; 125 undergraduates	HE; Ghana	RQ2, RQ3	Low	ChatGPT in a flipped classroom improved critical, reflective, and creative thinking compared with a control group.
15	Doshi & Hauser (2024)†	Exp	Online experiment; 293 writers, 600 evaluators	Adults; online	RQ3	High	AI-generated ideas raised individual creativity, especially among less creative writers, but reduced collective diversity.
16	Urban et al. (2024)	Exp	Preregistered experiment; 145 university students	HE; Czech Republic	RQ2, RQ3	High	ChatGPT improved elaboration, originality, and self-efficacy, but users misjudged the quality of their own work.
17	Habib et al. (2024)	Qual/Mixed	Mixed methods; creativity-course students	HE; USA	RQ3	Moderate	AI supports creative performance but may weaken creative self-confidence.
18	UNESCO (2024)	Concept	Competency framework	School students	RQ4	n/a	Four competency dimensions progressing through understand, apply, and create levels.
19	Ng et al. (2024)	Survey	Instrument validation; 363 students	Junior and SHS; Hong Kong	RQ4	Moderate	Validates the AILQ with affective, behavioral, cognitive, and ethical dimensions.
20	Ibanez (2024)	Survey	Descriptive-correlational; 166 students	Grades 7–12; Philippines	RQ2	Low	Moderate use of generative AI tools was reported alongside high self-reported critical thinking.
21	Kong et al. (2024)	Exp	Pre-post course evaluation with qualitative data; 128 students	SHS; Hong Kong	RQ4	Moderate	A 14-hour project-based course improved the application of AI concepts to real-world problems, metacognition, and understanding of AI ethics.
22	Bastani et al. (2025)	Exp	Randomized field experiment; nearly 1,000 students	SHS; Türkiye	RQ2, RQ5	High	Unguarded ChatGPT access lowered later scores by 17%; a guardrailed tutor improved practice by 127% and offset this harm.
23	Cukurova (2025)	Concept	Conceptual framework	Cross-level	RQ1	n/a	AI can externalize human thinking, be internalized by students, or extend cognition in integrated systems.

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

No	Study	Design category	Design and sample	Level and country	RQ	Quality	Key finding
24	Gerlich (2025)†	Survey	Survey with supplementary interviews; 666 respondents	General population	RQ2	Low	Frequent AI use was associated with cognitive offloading and weaker critical thinking, especially among younger users.
25	Zhang et al. (2025)	Survey	Instrument validation; 981 students	Middle school; USA	RQ4	Moderate	Validates the AI-CI concept inventory for middle school students.
26	Kong & Yang (2025)	Exp	Instrument validation and pre-post programme evaluation; 224 students	SHS and HE; Hong Kong	RQ4	Moderate	AI literacy attainment did not depend on prior programming experience.
27	Ning et al. (2025)	Survey	Instrument validation; 604 teachers	Teachers; China	RQ5	Moderate	Validates a four-dimension teacher AI literacy scale.
28	Sánchez Muñoz et al. (2025)	Qual/Mixed	Qualitative action research; high school students	SHS; Mexico	RQ1–RQ3	Moderate	Cycles of design thinking and prompt engineering supported creativity and computational fluency.
29	Mei et al. (2025)	Exp	Experiment; 225 university students	HE; United Kingdom	RQ3	Moderate	Generative AI boosted creative-writing performance but diminished the experience of creating.
30	Lhamo (2025)	Exp	One-group pre-experiment; 37 students	Grade 11; Bhutan	RQ3	Low	AI-generated feedback was followed by improved creative writing.
31	Sardi et al. (2025)	Review	Systematic review; 38 studies	Cross-level	RQ2	n/a	About 63% of studies reported gains in analytical and evaluative abilities.
32	Liu et al. (2025)	Review	Systematic review; 15 studies	EFL learning	RQ2	n/a	Ten of 15 studies reported positive effects under four design conditions.
33	Sari et al. (2025)	Survey	AI literacy survey; 542 university students	HE; Indonesia	RQ4	Moderate	AI literacy was relatively low in awareness, usage, evaluation, and ethics.
34	Pérez-Montesdeoca et al. (2026)	Review	Systematic review; 33 studies	Secondary school	RQ1–RQ5	n/a	Pedagogical integration of generative AI remains partial and short-term.
35	Wong & Qiu (2026)	Exp	Three-group experiment; 196 university students	HE	RQ1, RQ3	High	Thinking independently before using ChatGPT improved the transfer of independent creativity.
36	Nurcahyani & Noviani (2026)	Survey	Quantitative survey; 594 respondents	SHS and HE; Indonesia	RQ2	Moderate	Unstructured ChatGPT use was linked to dependence and weaker critical thinking.
37	Mahesh (2026)	Qual/Mixed	Exploratory qualitative; 19 students	Secondary school; Singapore	RQ1, RQ2	Moderate	The quality of Socratic AI dialogue depended on activity design and student readiness.
38	Putri & Ekasriadi (2026)	Exp	Quasi-experiment; 120 students	Grade 11 SHS; Indonesia	RQ2	Low	Critical literacy rose more in the AI-based experimental group than in the control group (d = 1.82).

Note. Exp = experimental or quasi-experimental (including pre-post evaluations); Survey = survey or instrument validation; Qual/Mixed = qualitative or mixed methods; Review = systematic or exploratory review; Concept = conceptual framework; HE = higher education; SHS = senior high school; n/a = MMAT not applicable (reviews and conceptual frameworks were not rated; see Methods). † General adult population, retained under the clarified criterion E2 and used as comparative evidence only.

3.3. Conceptualization and Operationalization of Human-AI Collaboration (RQ1)

The literature has moved from seeing AI as a tool to seeing it as a partner within a hybrid intelligence system, with control over learning remaining with students. Ouyang

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

and Jiao (2021) distinguish three paradigms: AI-directed, AI-supported, and AI-empowered, the last of which positions learners as leaders of their own learning. Cukurova (2025) extends this view with three conceptualizations: AI that externalizes human thinking, AI whose models students internalize, and integrated human-AI systems that extend cognitive capacity. According to Molenaar (2022), the greatest benefits come at intermediate levels of automation, where educators keep control.

On the regulation of learning, Järvelä et al. (2023) propose hybrid human-AI shared regulation that is triggered by critical moments in collaboration. Lodge et al. (2023) instead frame learning with generative AI as a network of co-regulation with individual agency at its center. Seo et al. (2021) found that AI systems can strengthen learner engagement but require clear allocation of responsibility and control.

At the operational level, the leading teachers in the study by Kim et al. (2022) recommended authentic, interdisciplinary, and creative tasks; process-based assessment; and instruction on AI principles and on analyzing AI errors. These teachers expected student-AI collaboration to develop in three stages: learning about AI, learning from AI, and learning with AI. Wong and Qiu (2026) demonstrated that activity sequence matters: students who generated their own ideas before engaging in dialogue with ChatGPT achieved greater transfer of independent creativity. Sánchez Muñoz et al. (2025) and Mahesh (2026) showed that, in secondary schools, the quality of collaboration depends on how activity cycles are designed and on whether students are ready to engage in meaningful dialogue.

3.4. Human-AI Collaboration and Critical Thinking (RQ2)

The evidence indicates that generative AI can enhance critical thinking when it is integrated through structured instructional design. Essel et al. (2024) reported, in a randomized experiment of low methodological quality, significant differences in critical, reflective, and creative thinking skills between undergraduate students who used ChatGPT in a flipped classroom and a control group. At the senior high school level in Indonesia, Putri and Ekasriadi (2026) reported, in a quasi-experiment of low methodological quality, that the experimental group's critical literacy rose from 56.4 to 84.7, compared with a rise from 55.9 to 67.2 in the control group; the effect size for the difference in gain scores ($d = 1.82$) was reported by the original study and was not recalculated in this review. The review by Sardi et al. (2025) found that about 63% of 38 studies reported gains in analytical and evaluative abilities, and Liu et al. (2025) found that 10 of 15 studies reported positive effects.

Design conditions determine whether these effects are positive or negative. Bastani et al. (2025) found that high school students who used a ChatGPT interface without guardrails scored 17% lower after access was withdrawn. A tutor version with pedagogical guardrails improved practice performance by 127% and largely offset these negative effects. Gerlich (2025) and Nurcahyani and Noviani (2026) linked unstructured use to cognitive offloading and dependence, particularly among younger users. Urban et al. (2024) added that AI users tend to misjudge the quality of their own work. Liu et al. (2025) identified four conditions for positive effects: critical evaluation of AI outputs, student-centered learning, AI used as scaffolding, and structured pedagogical integration.

3.5. Human-AI Collaboration and Creativity (RQ3)

The reviewed studies indicate that generative AI improves fluency, elaboration, and the quality of creative products, whereas originality and sense of ownership depend on how AI is involved. Most of this evidence comes from university students and adults. Urban et al. (2024) found gains in elaboration, originality, and self-efficacy in creative problem-solving among university students. In an online experiment with adult writers,

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

Doshi and Hauser (2024) showed that AI-generated ideas made individual stories more creative, especially among less creative writers. The same ideas also made the stories more similar to one another, which reduced collective diversity.

Habib et al. (2024) and Mei et al. (2025) found that AI supports creative performance but may weaken creative self-confidence and diminish the experience of creating. Wong and Qiu (2026) showed that a "think first, ChatGPT later" pattern led to better independent creativity on tasks done without AI. In secondary schools, the evidence is thinner: Sánchez Muñoz et al. (2025) reported growth in creativity and computational fluency through cycles of design thinking and prompt engineering in a qualitative action-research study, and Lhamo (2025) reported improved creative writing among Grade 11 students who received AI-generated feedback in a one-group pre-experiment of low methodological quality.

3.6. AI Literacy at the Secondary School Level (RQ4)

AI literacy is understood as a multidimensional construct that spans cognitive, usage, evaluative, affective, and ethical aspects. Long and Magerko (2020) set out AI literacy competencies under five themes, and Ng et al. (2021) identified four aspects. UNESCO (2024) defines four competency dimensions: a human-centered mindset, the ethics of AI, AI techniques and applications, and AI system design. These dimensions progress through three levels: understand, apply, and create.

Casal-Otero et al. (2023) found that very few of the learning experiences described in 179 documents assessed students' understanding after the intervention. Several validated instruments have begun to fill this gap: the AI-CI for middle school students (Zhang et al., 2025), the AILQ for secondary school students, which is built on affective, behavioral, cognitive, and ethical dimensions (Ng et al., 2024), and an AI empowerment instrument (Kong & Yang, 2025). A 14-hour project-based course improved senior high school students' ability to apply AI concepts to real-world problems, their metacognitive strategies, and their understanding of AI ethics (Kong et al., 2024). Kong and Yang (2025) also found that AI literacy attainment did not depend on prior programming experience. In the Indonesian context, Sari et al. (2025) reported that university students' AI literacy was still relatively low in awareness, usage, evaluation, and ethics.

3.7. Design Principles and Teacher Competence (RQ5)

Meaningful AI integration requires instructional materials that draw on teachers' technological, pedagogical, content, and ethical knowledge together. Celik (2023) showed that teachers' technological knowledge becomes meaningful only when combined with pedagogical knowledge. Mishra et al. (2023) argued that the generative and social nature of AI requires teachers to expand their contextual knowledge. Ning et al. (2025) validated a teacher AI literacy scale with four dimensions: perception, knowledge and skills, application and innovation, and ethics.

At the design level, Zawacki-Richter et al. (2019) noted that research on AI in education shows little pedagogical reflection, and Pérez-Montesdeoca et al. (2026) recommended integrated, sustained, and theory-based pedagogical models. Kasneci et al. (2023) argued that a pedagogical framework emphasizing critical thinking and fact-checking is a precondition for integrating large language models. Kim et al. (2022) added the need for flexible classrooms, a culture in which it is safe to fail, and school-level AI policies. Casal-Otero et al. (2023) recommended modular curricula developed together with educators.

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

3.8. Thematic Synthesis

Synthesizing across the research questions yielded four analytical themes. Together they link the benefits of AI for learning to the curricular designs that structure its use.

Theme 1: Pedagogical structure determines the direction of AI's benefits. Experiments with strong comparison groups, together with a systematic review, show that AI framed by pedagogical scaffolding improves performance and thinking skills, whereas unstructured access tends to create dependence (Bastani et al., 2025; Liu et al., 2025; Wong & Qiu, 2026). This positions instructional materials as a pedagogical guardrail that determines when, for what purposes, and in what ways students interact with AI. AI's contribution to critical thinking and creativity therefore depends on the curriculum design decisions that govern its use.

Theme 2: The sequence "humans initiate, AI extends, humans decide". Three findings show that the sequence of activities affects learning outcomes: the advantage of thinking independently before engaging with AI (Wong & Qiu, 2026), the problem of poorly calibrated self-assessment (Urban et al., 2024), and the loss of collective diversity in creative work (Doshi & Hauser, 2024). Students' initial ideas should come before AI outputs, AI outputs should be evaluated against explicit criteria, and final decisions about the work should stay with students. This pattern aligns with the learner-as-leader paradigm (Ouyang & Jiao, 2021) and with forms of hybrid regulation that preserve student agency (Järvelä et al., 2023; Lodge et al., 2023). Because these three studies involved university students and adults, the sequence is proposed for secondary schools as a design principle that still requires direct testing.

Theme 3: AI literacy as the link between critical thinking and creativity. Liu et al. (2025) identified the ability to write prompts and evaluate AI outputs as a prerequisite for positive effects on critical thinking. Kong et al. (2024) showed that a project-based AI literacy course also developed metacognitive and problem-solving strategies. These findings support the view that critical thinking, creativity, and AI literacy are interrelated. Understanding how AI works and where its limits lie makes critical evaluation possible, and critical evaluation in turn opens space for creative elaboration that goes beyond AI outputs.

Theme 4: Teachers and curriculum materials as determinants of implementation. AI integration requires teachers to combine technological, pedagogical, content, and ethical knowledge (Celik, 2023; Mishra et al., 2023). Yet a review at the secondary school level still found pedagogical integration to be partial and short-term (Pérez-Montesdeoca et al., 2026). This gap between conceptual frameworks and classroom practice shows a need for operational curriculum materials that specify objectives, activity sequences, ethical guidance, and assessment (Casal-Otero et al., 2023; Kim et al., 2022).

3.9. Comparison with Previous Reviews

The findings converge with earlier reviews in several respects and extend them in others. Like Sardi et al. (2025) and Liu et al. (2025), this review found that the effects of generative AI on critical thinking are predominantly positive but conditional on structured pedagogical integration and on students' evaluation of AI outputs. It extends these single-outcome reviews in two ways. First, the same conditionality was observed for creativity, where AI improved elaboration and product quality but could reduce originality, ownership, and collective diversity (Doshi & Hauser, 2024; Mei et al., 2025). Second, activity sequence emerged as a design variable in its own right (Wong & Qiu, 2026), which the earlier reviews did not identify.

The review also diverges from earlier syntheses. The largely positive picture reported by Sardi et al. (2025) contrasts with the negative effects of unguarded access

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

found in the randomized field experiment of Bastani et al. (2025). This review reconciles the two by treating pedagogical structure as the condition that determines the direction of AI's effects (Theme 1) rather than treating AI use as beneficial or harmful in itself. With regard to AI literacy, Casal-Otero et al. (2023) noted that K-12 interventions rarely assessed learning outcomes. The present review shows that this gap has been partly addressed since 2023 through validated instruments for secondary school students (Kong & Yang, 2025; Ng et al., 2024; Zhang et al., 2025), although none has yet been validated with Indonesian senior high school students. Consistent with Pérez-Montesdeoca et al. (2026), integration in secondary schools was found to be partial and short-term; this review goes further by specifying the sequence of activities and the evidence status of each phase that such integration could follow.

Taken together, the four analytical themes make three theoretical contributions. They operationalize the AI-empowered paradigm (Ouyang & Jiao, 2021) and the concept of hybrid intelligence (Molenaar, 2022) at the level of classroom activity sequences, in the form of the principle that humans initiate, AI extends, and humans decide. They position AI literacy as the competency that links critical evaluation and creative elaboration, rather than as a parallel learning outcome. Finally, they extend the notion of guardrails from the design of AI tools (Bastani et al., 2025) to the design of curriculum materials, which can regulate when and how students interact with AI even when the tool itself has no built-in safeguards.

3.10. Evidence Review of the Five-Phase Framework (RQ6)

The five-phase framework for human-AI collaborative learning (Human-AI Collaborative Learning, HAICL) was formulated in the authors' development research before this review was conducted (Figure 3). In Figure 3, the solid arrows show the order of the five phases within one learning cycle, and the dashed lines to the right link each phase to the competencies it targets. The dashed arrow from Reflection back to Exploration represents the iterative nature of the framework: the results of reflection, such as which AI outputs required closer verification or how prompts should be reformulated, become the starting point for exploration in the next learning cycle. It does not denote an additional step within a single lesson. Table 6 maps each phase as formulated in the framework against the synthesized evidence and shows its level of support and the confidence in that classification.

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

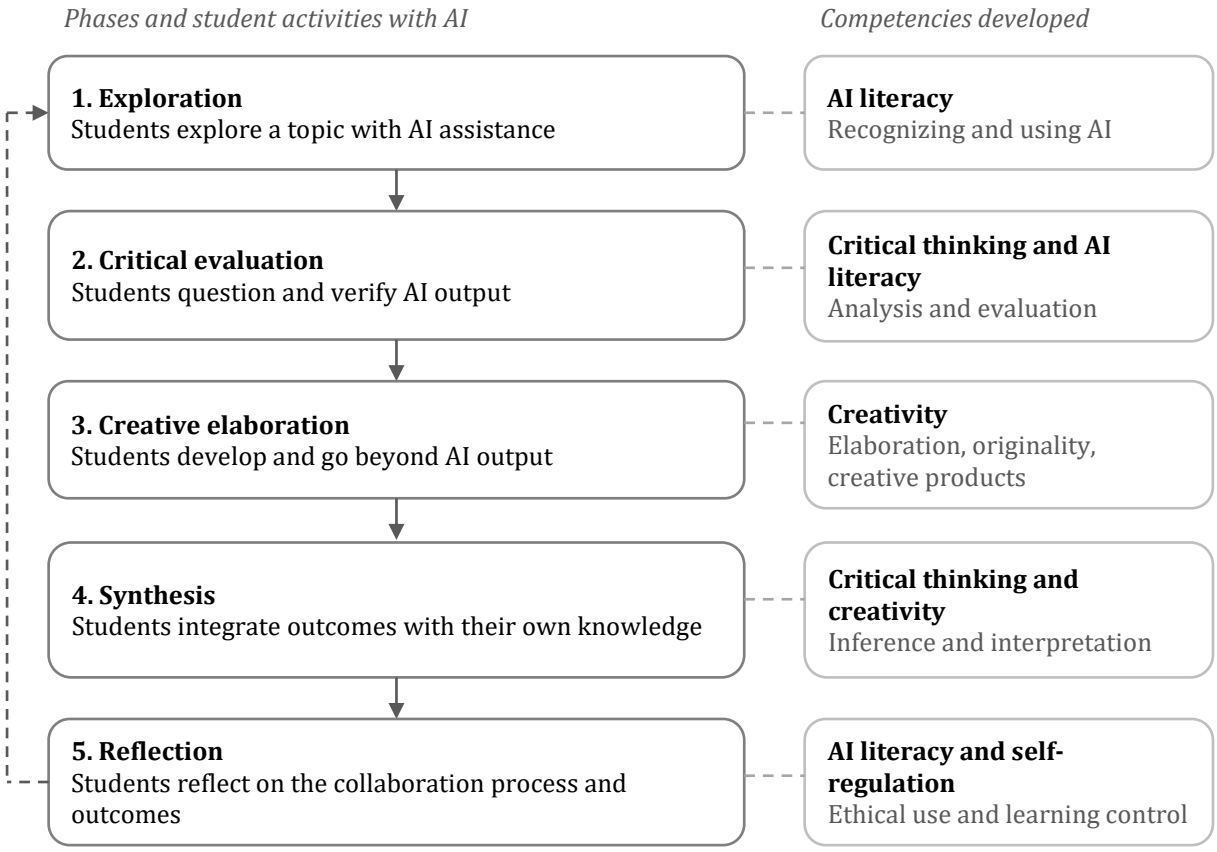


Figure 3. Five-phase human-AI collaborative learning framework and target competencies.

Within the framework, the reflection phase is intended to build metacognitive awareness and to improve the quality of human-AI collaboration in subsequent learning cycles.

Table 6. Evidence review for each phase of the framework.

Phase	Formulation in the framework	Synthesized evidence	Level of support	Confidence and relevance to secondary schools
Exploration	Students explore a topic with AI assistance to build initial understanding and identify multiple perspectives	AI broadens the range of ideas (Doshi & Hauser, 2024; Urban et al., 2024), but generating initial ideas independently before using AI leads to better skill transfer (Wong & Qiu, 2026), and unstructured AI access lowers independent learning outcomes (Bastani et al., 2025)	Partially supported	Moderate: coherent high-quality evidence, including one secondary school experiment; the sequence finding comes from higher education

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

Phase	Formulation in the framework	Synthesized evidence	Level of support	Confidence and relevance to secondary schools
Critical evaluation	Students evaluate, question, and verify AI outputs	Pedagogical scaffolding reduces dependence among high school students (Bastani et al., 2025); AI-based critical literacy tasks improve senior high school students' critical literacy (Putri & Ekasriadi, 2026); evaluating AI outputs is the key condition for positive effects on critical thinking (Liu et al., 2025; Sardi et al., 2025); explicit criteria are needed because self-assessment is poorly calibrated (Urban et al., 2024)	Supported by empirical evidence	Moderate: one high-quality randomized experiment in senior high schools and one high-quality university experiment, consistent with two reviews; the second secondary school study is of low quality
Creative elaboration	Students develop AI outputs and go beyond them to produce more original work	AI improves elaboration and product quality (Urban et al., 2024), while personal contributions preserve the originality, ownership, and diversity of the work (Doshi & Hauser, 2024; Habib et al., 2024; Mei et al., 2025); secondary school evidence is limited to one qualitative study and one low-quality pre-experiment (Lhamo, 2025; Sánchez Muñoz et al., 2025)	Supported by empirical evidence (higher education and adult samples)	Low for secondary schools: high-quality evidence, but downgraded for relevance because it comes mainly from university students and adults; transferability not yet established

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

Phase	Formulation in the framework	Synthesized evidence	Level of support	Confidence and relevance to secondary schools
Synthesis	Students integrate the outcomes of collaboration with their own thinking and knowledge	Consistent with the concept of hybrid intelligence, which emphasizes students' internalization of AI models while control remains with humans (Cukurova, 2025; Molenaar, 2022), and with project-based learning that combines AI concepts with real-world problem-solving (Kong et al., 2024)	Supported by conceptual evidence	Low: no study tested the phase on its own
Reflection	Students reflect on the process and outcomes of their collaboration with AI to build metacognitive awareness	Regulation and co-regulation preserve student agency (Järvelä et al., 2023; Lodge et al., 2023); excessive cognitive offloading needs to be curbed through awareness of AI's role (Gerlich, 2025)	Supported by conceptual evidence	Low: conceptual frameworks and one adult survey; no study tested the phase on its own

Under the decision rules, the critical evaluation phase has the strongest support, because it rests on experiments with comparison groups, including a high-quality randomized field experiment and a lower-quality quasi-experiment, both conducted in senior high schools (Bastani et al., 2025; Putri & Ekasriadi, 2026) and is consistent with two systematic reviews. The requirement of two studies of at least moderate quality is met by the field experiment and by the high-quality university experiment on poorly calibrated self-assessment (Table 6), not by the quasi-experiment. The creative elaboration phase also meets the rule for empirical support, but the contributing experiments involved university students (Mei et al., 2025; Urban et al., 2024; Wong & Qiu, 2026) or adult writers (Doshi & Hauser, 2024). For secondary schools, this classification should therefore be read as provisional: confidence was rated low, and the phase requires direct testing with secondary school students before it can be regarded as empirically supported in that context. The synthesis and reflection phases are consistent with frameworks of hybrid intelligence and co-regulation, but no empirical study was found that tested either phase on its own.

The exploration phase is partially supported. Starting exploration with AI assistance fits the finding that AI broadens the range of ideas. It conflicts, however, with evidence that generating ideas independently before using AI builds skills that last once AI is no longer available (Bastani et al., 2025; Wong & Qiu, 2026). This review therefore recommends adding a short step in which students formulate initial ideas without AI before they interact with it, while keeping AI's role in the exploration phase. Table 7

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

illustrates how one learning cycle in the KKA subject could be organized with the revised exploration phase.

Table 7. Example syntax of one KKA learning cycle with the revised exploration phase.
Topic: Digital content creation through human-AI collaboration (KKA, Grade XI); one meeting of 3 × 45 minutes

Phase	Time (min)	Student activity	Teacher role	Output
1a. Independent idea generation (no AI)	15	Students individually draft the purpose, audience, and key message of a digital content product on a worksheet.	Presents the problem and the success criteria; ensures that AI is not used yet.	Initial idea sheet
1b. AI-assisted exploration	20	Students write prompts with five elements (role, task, context and audience, output format, constraints) to broaden alternatives, then compare AI suggestions with their own ideas.	Models prompt writing and asks students to note which ideas are their own.	Prompt log
2. Critical evaluation	25	Students verify AI suggestions against a checklist (accuracy, relevance, bias, copyright and ethics) and cross-check claims with at least two other sources.	Provides explicit criteria and demonstrates verification of one AI output.	Completed evaluation checklist
3. Creative elaboration	35	Students develop the product by combining their own ideas with verified AI elements and record their personal contributions.	Prompts students to go beyond AI outputs and to justify their design choices.	Draft product and contribution log
4. Synthesis	20	Groups present the product and explain which decisions were made by students and which drew on AI.	Facilitates discussion and links the product to the learning objectives.	Final product with justification

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

Phase	Time (min)	Student activity	Teacher role	Output
5. Reflection	20	Students write a reflection on AI's role, what they verified, signs of over-reliance, and ethical issues, and set a goal for the next cycle.	Gives feedback on the reflections and uses them to plan the next cycle.	Reflection journal (input for the next cycle)

3.11. Practical Implications

For teachers, these findings suggest that classroom AI use should start from students' own ideas and that every AI output should come with a verification task based on explicit criteria. For curriculum developers and teaching module designers, particularly in the KKA subject, the five phases, with the proposed change to the exploration phase, can serve as a framework for organizing activity flows, worksheets, and assessments that cover both process and product, as illustrated in Table 7. For schools, the findings on teacher competence point to the need for professional development that combines technological, pedagogical, and ethical knowledge, and for school-level policies on AI use.

3.12. Research Gaps

The synthesis identified four research gaps. First, experimental studies with comparison groups at the senior high school level are still scarce. Of the 10 experimental studies, only Bastani et al. (2025) and Putri and Ekasriadi (2026) were conducted in senior high schools with comparison groups. Second, no study measured critical thinking, creativity, and AI literacy at the same time. Third, no study was found that developed and tested human-AI collaborative instructional materials for the KKA subject. Fourth, evidence from Indonesia comes mostly from surveys and has not yet explicitly tested the phases of human-AI collaboration.

4. CONCLUSION

This integrative review of 38 studies concludes that generative AI can strengthen students' critical thinking, creativity, and AI literacy when it is integrated through instructional designs that are structured and sequenced and that keep students in control of their learning. Human-AI collaboration is conceptualized as hybrid intelligence and co-regulation (RQ1). Its effects on critical thinking depend on pedagogical scaffolding and on tasks that require students to evaluate AI outputs (RQ2). AI improves elaboration and the quality of creative products, and originality is preserved when students think independently before using AI, although this evidence comes mainly from higher education and adults (RQ3). AI literacy is a multidimensional construct that develops well through project-based learning, and validated instruments for secondary schools now exist (RQ4). Implementation depends on teachers combining technological, pedagogical, and ethical knowledge and on the availability of operational curriculum materials (RQ5). In the five-phase framework, the critical evaluation phase is supported by empirical evidence with moderate confidence. The creative elaboration phase is supported by empirical evidence from higher education and adult samples, and its transferability to secondary schools remains to be tested. The synthesis and reflection phases are supported mainly by conceptual work, and the exploration phase is partially supported and should begin with students formulating their own ideas before interacting with AI (RQ6).

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

This review has several methodological limitations. First, study identification was not fully reproducible. The included studies came from a thesis reference library built through topic-based searches in Consensus and Google Scholar and from supplementary Google Scholar searches and AI tool recommendations that were not logged. A post hoc structured search retrieved none of the included studies, so relevant studies may have been missed and the selection may favor sources that these tools rank highly. Second, the numbers of candidate records assessed and excluded were not recorded. Third, a single reviewer selected the studies and appraised their quality, and inter-rater agreement could not be calculated; preliminary quality ratings were drafted with AI assistance before being checked by the first author. These conditions increase the risk of selection and appraisal bias, particularly because the five-phase framework was formulated by the authors before this review was conducted. The risk was reduced, but not eliminated, through pre-specified eligibility criteria, verification of every AI-recommended study on the publisher's website, operational decision rules for classifying evidence, and the explicit reporting of evidence that did not fit the framework. Fourth, most of the experimental evidence comes from higher education and adult samples, which limits transferability to secondary schools, and because preprints were excluded, some recent findings may be missing. The findings should therefore be read as an evidence-informed basis for instructional design rather than an exhaustive synthesis. A future systematic review should search Scopus, Web of Science, and ERIC with logged search strategies and use dual independent screening with a calculated measure of inter-rater agreement.

Future research should conduct controlled experiments at the senior high school level that measure critical thinking, creativity, and AI literacy at the same time; test how the sequence of human-AI collaboration phases affects students' transfer of independent skills; test the creative elaboration phase directly with secondary school students; adapt and validate AI literacy instruments for senior high school students in the Indonesian context; and test the refined five-phase framework for validity, practicality, and effectiveness by developing Coding and Artificial Intelligence (KKA) instructional materials, including empirical tests of the synthesis and reflection phases.

5. AUTHORS' NOTE

The authors declare that they have no competing financial interests or personal relationships that could have influenced the work reported in this manuscript, and that the research and the writing of the manuscript adhered to the standards of research and publication ethics and are free from plagiarism.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability. The annotated PRISMA 2020 checklist (Supplementary Material S1); the sources and provenance of each included study, the post hoc Publish or Perish search results, and the data extraction table (Supplementary Material S2); and the criterion-level MMAT ratings (Supplementary Material S3) are provided as supplementary material.

Use of generative AI. Generative artificial intelligence tools were used in three ways. First, AI-assisted literature tools (Consensus and general-purpose AI assistants) were used to identify candidate studies; every recommended study was verified on the publisher's website and read in full before inclusion. Second, a generative AI tool was used to draft preliminary MMAT ratings, which the first author checked and finalized against the full texts. Third, Claude AI assisted with brainstorming ideas, discussing

concepts, outlining, and drafting or refining text, while Grammarly was used to correct grammar and spelling, edit and paraphrase the language, and improve the clarity, readability, style, and coherence of the writing. AI tools did not make inclusion decisions or extract data. The authors reviewed and edited all AI-assisted content to ensure its accuracy, completeness, and compliance with ethical and scholarly standards, and they take full responsibility for the final version of the manuscript. This statement follows the journal's policy on the use of AI in manuscript preparation.

6. REFERENCES

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

BSKAP. (2025). *Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Nomor 046/H/KR/2025 tentang Capaian Pembelajaran pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah pada Kurikulum Merdeka*. Kementerian Pendidikan Dasar dan Menengah Republik Indonesia.

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education, 10*(1), 29. https://doi.org/10.1186/s40594-023-00418-7

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

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology, 56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances, 10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity, 34*(1), 100072. https://doi.org/10.1016/j.yjoc.2023.100072

Harzing, A. W. (2007). *Publish or Perish* [Computer software]. https://harzing.com/resources/publish-or-perish

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Ibanez, J. (2024). Effects of utilization of generative artificial intelligence tools on critical thinking skills of high school students. *Proceedings of the International Scholars Conference, 11*(4), 1190–1201. https://doi.org/10.35974/isc.v11i4.3601

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology, 54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Kong, S.-C., Cheung, M.-Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence, 6*, 100214. https://doi.org/10.1016/j.caeai.2024.100214

Kong, S.-C., & Yang, Y. (2025). Developing and validating an artificial intelligent empowerment instrument: Evaluating the impact of an artificial intelligent literacy programme for secondary school and university students. *Research and Practice in Technology Enhanced Learning, 20*, 024. https://doi.org/10.58459/rptel.2025.20024

Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M. A., Tunçalp, Ö., Colvin, C. J., Garside, R., Carlsen, B., Langlois, E. V., & Noyes, J. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings: Introduction to the series. *Implementation Science, 13*(Suppl. 1), 2. https://doi.org/10.1186/s13012-017-0688-3

Lhamo, D. (2025). ChatGPT: An enabling model for English creative writing in the eleventh grade of Phuentsholing Rigsar Higher Secondary School. *Zorig Melong: A Technical Journal of Science, Engineering and Technology, 8*(2), 128–133. https://doi.org/10.17102/zmv8.i2.013

Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments, 12*(1), 48. https://doi.org/10.1186/s40561-025-00406-0

Lodge, J. M., de Barba, P., & Broadbent, J. (2023). Learning with generative artificial intelligence within a network of co-regulation. *Journal of University Teaching and Learning Practice, 20*(7). https://doi.org/10.53761/1.20.7.02

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

Mahesh, K. S. (2026). Can AI teach like Socrates? An exploratory study of AI-student dialogue in secondary physics education. In *The Asian Conference on Education 2025: Official Conference Proceedings* (pp. 1677–1691). The International Academic Forum. https://doi.org/10.22492/issn.2186-5892.2026.127

Mei, P., Brewis, D. N., Nwaiwu, F., Sumanathilaka, D., Alva-Manchego, F., & Demaree-Cotton, J. (2025). If ChatGPT can do it, where is my creativity? Generative AI boosts

performance but diminishes experience in creative writing. *Computers in Human Behavior: Artificial Humans, 4*, 100140. https://doi.org/10.1016/j.chbah.2025.100140

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4), 632–645. https://doi.org/10.1111/ejed.12527

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*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). 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

Ning, Y., Zhang, W., Yao, D., Fang, B., Xu, B., & Wijaya, T. T. (2025). Development and validation of the Artificial Intelligence Literacy Scale for Teachers (AILST). *Education and Information Technologies, 30*(12), 17769–17803. https://doi.org/10.1007/s10639-025-13347-5

Nurcahyani, S. T., & Noviani, L. (2026). Pengaruh penggunaan ChatGPT terhadap kemampuan berpikir kritis pada siswa SMA Negeri 1 Kartasura dan mahasiswa S-1 Universitas Sebelas Maret. *J-KIP (Jurnal Keguruan dan Ilmu Pendidikan), 7*(2), 686–702. https://doi.org/10.25157/j-kip.v7i2.20097

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence, 2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

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*, n71. https://doi.org/10.1136/bmj.n71

Pérez-Montesdeoca, H., Rodriguez-Rodriguez, D., & Fernández-Sogorb, A. (2026). The integration of generative artificial intelligence in secondary education: A systematic review. *Contemporary Educational Technology, 18*(3), ep667. https://doi.org/10.30935/cedtech/18746

Putri, N. K. S. S., & Ekasriadi, I. A. A. (2026). Strengthening AI-based critical literacy in literature learning under the Merdeka Curriculum. *Jurnal Konfiks, 13*(2), 53–62. https://doi.org/10.26618/tt4n9b68

Sánchez Muñoz, J. A., Flores-Eraña, G., Silva-Campos, J. M., Chavira-Quintero, R., & Olais-Govea, J. M. (2025). GenAI as a cognitive mediator: A critical-constructivist inquiry into computational thinking in pre-university education. *Frontiers in Education, 10*, 1597249. https://doi.org/10.3389/feduc.2025.1597249

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy (iJEP), 15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Sari, D. K., Supahar, S., Rosana, D., Dinata, P. A. C., & Istiqlal, M. (2025). Measuring artificial intelligence literacy: The perspective of Indonesian higher education students. *Journal of Pedagogical Research, 9*(2), 143–157. https://doi.org/10.33902/JPR.202531879

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(1), 54. https://doi.org/10.1186/s41239-021-00292-9

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

UNESCO. (2024). *AI competency framework for students*. UNESCO. https://doi.org/10.54675/JKJB9835

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Wong, S. S. H., & Qiu, S. X. (2026). Think first, ChatGPT later: Guiding human–AI collaboration for learning gains in independent human creativity. *Educational Psychology Review, 38*(1), 45. https://doi.org/10.1007/s10648-026-10118-7

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, H., Perry, A., & Lee, I. (2025). Developing and validating the Artificial Intelligence Literacy Concept Inventory: An instrument to assess artificial intelligence literacy among middle school students. *International Journal of Artificial Intelligence in Education, 35*(1), 398–438. https://doi.org/10.1007/s40593-024-00398-x

DOI: https://doi.org/10.17509/e.v25i3.xxxxxx
P-ISSN 0852-1190 E-ISSN 2502-0781