**Journal of Current Studies in SDGs**
Homepage : https://journal.i-ros.org/index.php/JOCSIS
Email : jocsis@i-ros.org

e-ISSN 3089-7564
Vol.4, No.1, March 2027
© 2025 Journal of Current Studies in SDGs

# Generative AI-Supported Learning in Pre-Service Science Teacher Education: A Systematic Review of Experiential Learning and Reflective Thinking for SDG 4

**An Nuril Maulida Fauziah, Tutut Nurita, Ahmad Fauzi Hendratmoko, Martini, Enny Susiyawati, Dyah Astriani, Erman, Fatimatuz Zahro**

State University of Surabaya, Surabaya, Indonesia

Check for updates OPEN ACCESS

**DOI : https://doi.org/10.63230/jocsis.4.1.347**

**Sections Info**

***Article history:***
Submitted: July 25, 2026
Final Revised:September 26, 2026
Accepted: October 2, 2026
Final Available Online: November 13, 2026
Published Date: March 27, 2028

***Keywords:***
Experiental Learning;
Generative Artificial;
Intelligence;
Kolb's Learning:
Pre-service Science Teacher;
Science Education;
SDG 4.

**ABSTRACT**

***Objective:*** *The review synthesises how generative artificial intelligence (GenAI) is used in pre-service teacher education, with particular attention to science and STEM contexts, experiential learning alignment, reflective thinking, and the implications for Sustainable Development Goal 4 (SDG 4).* ***Method:*** *The review screened 411 unique records exported from Scopus and identified a priority subset of 45 records. In the updated working assessment, 35 records had received detailed assessment: 30 full texts and five abstract/metadata-only records. Twenty-two full-text records were retained for supporting or contextual synthesis, eight were excluded, three remained unclear, and two were provisionally retained on abstract evidence. Ten priority records remained pending assessment.* ***Results:*** *GenAI was used as a dialogic reflection partner, lesson-design assistant, cognitive or metacognitive scaffold, research-learning support, and interactive tutor or simulation agent. However, the studies varied in population, disciplinary context, theoretical framework, and outcome measurement. Experiential and reflective activities were reported, but explicit implementation of all four stages of Kolb's cycle was not established in the studies assessed. Direct reflective-thinking measures were limited and were mainly reported in studies outside science/STEM teacher education. No assessed study simultaneously combined a clearly implemented four-stage Kolb cycle, direct reflective-thinking measurement, and a pre-service science/STEM teacher population.* ***Novelty:*** *The review distinguishes GenAI functions, experiential-learning alignment, and different forms of reflection-related evidence. It identifies the need for theoretically structured, human-centred GenAI-supported learning designs that can strengthen reflective thinking in science teacher education and contribute to SDG 4.*

## INTRODUCTION

Science teacher education requires prospective teachers to integrate disciplinary understanding with inquiry-oriented pedagogy, evidence-based reasoning, classroom decision-making, and the capacity to respond to diverse learners (Breslyn & McGinnis, 2019). These demands are not met through content acquisition alone. Pre-service teachers also need opportunities to examine their instructional decisions, interpret evidence from teaching and learning, identify assumptions, and revise pedagogical approaches (Hordvik et al., 2024). Reflective thinking is therefore relevant not merely as a retrospective activity, but as a professional reasoning process through which experience can be evaluated and transformed into informed action (Sherwood, 2024).

The rapid development of generative artificial intelligence (GenAI) has introduced new possibilities for supporting such processes. Conversational large language models can respond to prompts, offer alternative explanations, generate questions, critique lesson plans, and provide iterative feedback (Beale, 2025). In teacher education, these affordances may position GenAI as a dialogic scaffold or a cognitive partner rather than only as a content-production tool (Spasopoulos et al., 2025). Recent studies have

explored ChatGPT-supported lesson planning, reflective writing, collaborative lesson study, and simulated teaching interactions. For example, Abualrob (2025) examined ChatGPT-assisted re-reflection during an elementary science teaching practicum, while Katsampoxaki-Hodgetts (2026) investigated how pre-service science teachers used several GenAI tools as dialogic partners during lesson redesign (Abualrob, 2025; Katsampoxaki-Hodgetts, 2026; Zhang et al., 2026). Zhang et al. (2026) explored science and mathematics pre-service teachers' experiences in a GenAI-powered virtual-reality teaching simulation.

These developments should not, however, be interpreted as evidence that GenAI automatically improves reflective thinking. The educational value of a chatbot depends on the task, the quality of prompts and feedback, learners' capacity to evaluate AI outputs, and the extent to which the activity preserves human judgement (Chang et al., 2023). Reflection may be documented through journals, dialogue, lesson revisions, or interaction logs; these forms of evidence are not interchangeable with a direct measure of reflective-thinking ability. A study that analyses reflective journals, for example, may provide evidence about reflective practice without demonstrating a measured change in a reflective-thinking construct. Similarly, metacognitive self-regulation, critical thinking, epistemic reflexivity, and reflective thinking are related but distinct constructs and should be reported separately (Dinsmore et al., 2008).

Experiential learning provides a useful pedagogical frame for connecting experience with reflection and subsequent action. In Kolb's experiential learning theory, learning is conceptualised as a cyclical transformation of experience through four modes: concrete experience, reflective observation, abstract conceptualisation, and active experimentation (Kolb, 2015). These stages offer a structured way to design learning activities in which pre-service teachers engage with a science or STEM problem, reflect on what occurred, connect observations with disciplinary concepts, and test revised explanations or designs. A lab investigation, field-based inquiry, engineering design task, or microteaching episode may provide a concrete experience; structured questioning and evidence review may support reflective observation; scientific explanation and theory-building may support abstract conceptualisation; and redesign or re-testing may enact active experimentation. A study should be classified as Kolb-based only when its design substantively addresses the stages, not simply because it includes an experience or a reflection activity (Kolb, 2015).

Within such a cycle, a GenAI chatbot could provide carefully bounded scaffolding. During concrete experience, it may prompt learners to notice phenomena or record initial predictions. During reflective observation, it may ask learners to compare predictions with observations, identify anomalies, or justify interpretations. During abstract conceptualisation, it may help learners connect evidence to scientific models while requiring verification against authoritative sources. During active experimentation, it may support the planning of a revised procedure, prototype, or teaching strategy. This potential depends on instructional design: AI-generated responses need to be scrutinised, and learners should remain responsible for evaluating evidence and making decisions (Krushinskaia et al., 2026).

The relevance of this agenda extends to Sustainable Development Goal 4 (SDG 4), which calls for inclusive and equitable quality education and lifelong learning opportunities for all. In teacher education, SDG 4 is particularly relevant to the preparation of qualified teachers, the development of relevant skills, and education for sustainable development. These connections provide a policy-oriented interpretive lens for considering how responsible GenAI integration might support teacher learning.

They do not mean that a study has empirically measured SDG 4 attainment. In this review, SDG 4 is therefore treated as a framework for interpreting educational implications rather than as an outcome variable unless an individual study explicitly operationalises it (Johnston, 2016).

The working evidence map indicates that the relevant research is dispersed across different disciplines and theoretical traditions. Some studies focus on pre-service science teachers and GenAI-mediated lesson design; others examine reflective thinking in general teacher education, mathematics education, English language teacher education, or broader professional learning. The studies also differ in their treatment of experiential learning: some mention Kolb theoretically, some use another reflective or design framework, and others do not identify an experiential-learning model. Direct measures of reflective thinking are reported in some studies outside the specific science-teacher context, whereas several science-related studies examine reflection qualitatively through journals, interaction traces, interviews, or lesson redesign. This heterogeneity makes a critical synthesis necessary (Campbell et al., 2020).

Accordingly, this review maps the available evidence without treating all forms of AI use, experience-based learning, or reflective practice as equivalent. Its contribution is to clarify what has been investigated, distinguish direct evidence from supporting evidence, examine how experiential learning and reflective thinking are operationalised, and identify the gap in integrated designs for pre-service science teachers. The review also considers how future designs might align with SDG 4 through inclusive, responsible, and professionally meaningful science teacher preparation.

Review Questions

- RQ1. How are GenAI chatbots and related GenAI tools used to support learning and reflective thinking among pre-service teachers, particularly in science and STEM education?
- RQ2. To what extent are the learning activities reported in the reviewed studies aligned with experiential learning theory, particularly the four stages of Kolb's experiential learning cycle?
- RQ3. What evidence is reported regarding reflective-thinking development and its assessment, what implementation challenges are identified, and what implications can be drawn for quality science teacher education in relation to SDG 4?

## RESEARCH METHOD

### *Review design and reporting approach*

This review adopted a systematic literature review design with a critical narrative and thematic synthesis. The review examines the intersection of GenAI-supported learning, experiential learning, reflective thinking, and pre-service science teacher education. Reporting is guided by PRISMA 2020. PRISMA is treated as a reporting guideline and does not replace a reproducible search strategy, transparent eligibility criteria, complete screening, or methodological quality appraisal.

### *Information source and search record*

The initial database comprised 411 unique records exported from Scopus and organised into nine screening batches covering article numbers 1–411. The database contains bibliographic metadata, abstracts, DOI information, title–abstract decisions, rationales, and batch identifiers. The exact Scopus search string, search date, publication-year limits, and database-interface filters are not recorded in the current manuscript copy. These details must be recovered from the original Scopus search history and reported

verbatim; they should not be reconstructed from memory or inferred from the retrieved records.

### *Eligibility principles*

Eligibility decisions were guided by five dimensions: population, participant interaction with GenAI, disciplinary context, experiential-learning framework, and reflective-thinking construct. Evidence from pre-service science teachers in science/STEM settings was treated as most directly relevant. Studies involving other teacher-education populations or disciplines were retained only as supporting or contextual evidence when they informed GenAI-supported learning, reflection, or teacher readiness. The review distinguished direct learner interaction with GenAI from studies that only analysed AI-generated artefacts or examined intentions to use AI.

**Table 1.** Eligibility criteria and operational definitions for study selection

| Eligibility Dimension | Operational Rule |
|---|---|
| Population | Core evidence: pre-service science teachers. Pre-service teachers from other disciplines, in-service teachers, and general teacher-education samples may be retained as supporting or contextual evidence, but are not counted as direct evidence for the target population. |
| GenAI role | Core relevance requires GenAI/chatbot interaction to be used directly by participants as part of learning, reflection, instructional design, or pedagogical reasoning. AI used only by researchers for data analysis, or only by teaching staff to generate materials without learner interaction, is not treated as a chatbot-supported learner intervention. |
| Disciplinary context | Science and STEM education are the core contexts. Mathematics, language education, general teacher education, and other fields may contribute supporting evidence where the GenAI and reflective-learning mechanisms are relevant. |
| Experiential learning | Experiential learning is not equated with any activity described as experiential or reflective. Kolb-based evidence requires examination of concrete experience, reflective observation, abstract conceptualisation, and active experimentation. Alternative models (e.g., ALACT, Schön, Learning by Design, 5E, or lesson study) are recorded as distinct frameworks. |
| Reflective construct | Direct reflective-thinking measurement is distinguished from qualitative reflective practice, reflective journals, metacognition, self-regulated learning, critical thinking, and epistemic reflexivity. The construct and measurement approach are recorded separately. |
| Study type | Empirical studies are prioritised for the evidence synthesis. Reviews, conceptual papers, protocols, and studies where AI is used only for automated analysis are not treated as equivalent to empirical chatbot-supported learning interventions. |

### *Screening and progress of assessment*

The 411 records underwent title–abstract screening across nine batches, resulting in 49 Include and 163 Exclude decisions, while 199 records were selected for further consideration because eligibility could not be confirmed from titles and abstracts alone. A priority subset of 45 records was selected for detailed follow-up, and the rationale for its selection should be stated explicitly in the final methods. The updated working assessment covers 35 priority records: 30 full texts were examined, with 22 retained for supporting or contextual synthesis and eight excluded; five records remain abstract/metadata-only (three unclear and two provisionally retained as supporting

evidence), while ten remain pending. Four full-text assessments completed after database version 29 (articles 26, 41, 42, and 44) are reflected in this synthesis; article 42 is used only as contextual evidence on adoption intention and readiness, not as an implemented learning intervention. These figures remain provisional and must be reconciled with the master screening workbook before finalising the PRISMA flow diagram, which summarises the selection and screening process (Figure 1).

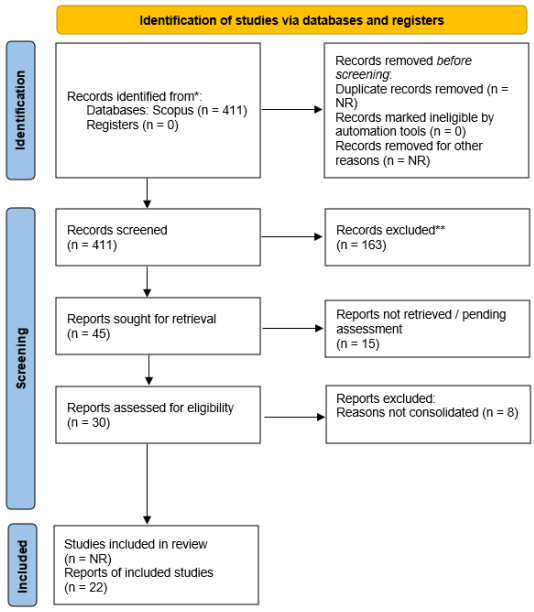


**Figure 1.** PRISMA 2020 flow diagram of the study selection process.

### *Data extraction and evidence classification*

For each detailed record, the review captured article identity, study type, participant group, disciplinary context, GenAI tool and role, experiential-learning framework or stages, reflective-thinking construct and measurement, study design and sample, key findings, limitations, and reviewer comments. Evidence was classified along two dimensions: (a) direct science/STEM teacher-education evidence versus supporting or contextual evidence from other populations or disciplines; and (b) full-text assessment versus provisional abstract/metadata assessment. Full-text exclusions were retained with their reasons to preserve an auditable decision trail.

### *AI-assisted support in the review workflow*

ChatGPT (OpenAI) was used as an auxiliary tool during the development of this review to support the summarisation of selected abstracts and accessible full-text articles, organisation of article-level extraction notes, preparation of preliminary screening rationales, and structuring of manuscript drafts. AI-generated summaries, classifications, and draft passages were treated as provisional working material rather than primary evidence. The review team remains responsible for final eligibility decisions and interpretation. Before submission, every bibliographic detail, extracted result, numerical value, and interpretive claim must be checked against the original Scopus record or article. The team should retain prompts, dates of use, outputs adopted or rejected, and records of human verification in accordance with the journal's disclosure requirements.

### *Synthesis strategy*

A narrative and thematic synthesis was used because the reviewed studies differ in research design, participant group, GenAI implementation, learning framework, and reflection-related outcome. The synthesis is organised directly around the three research questions: (1) GenAI applications in teacher learning (RQ1); (2) alignment of reported learning activities with experiential learning theory and Kolb's four stages (RQ2); and (3) evidence and assessment of reflective thinking, implementation challenges, and implications for SDG 4 (RQ3). Quantitative findings are reported only when stated in the source and are not pooled across different constructs or instruments.

### *Quality appraisal and review limitations*

The working database contains study-level methodological notes and selected reliability or coding information, but a single standardised quality-appraisal instrument has not yet been applied consistently across all retained studies. Accordingly, this draft does not rank the overall certainty of evidence or pool effect estimates. A suitable appraisal of qualitative, quantitative, and mixed-methods studies, verification of the Scopus search strategy and date, justification of the 45-record priority subset, assessment of the ten pending records, and completion of the PRISMA flow are required before submission.

## RESULTS AND DISCUSSION

### *Results*

### *Characteristics of the reviewed evidence*

The updated working synthesis draws on research concerning GenAI-supported learning, instructional design, reflection, and teacher education, with particular attention to science and STEM contexts. Of the 45 priority records, 35 have received detailed assessment. These comprise 22 full-text records retained for supporting or contextual synthesis, eight full-text exclusions, three abstract/metadata-only records classified as unclear, and two abstract-only records provisionally retained as supporting evidence. Ten priority records remain pending. The present profile is therefore a working synthesis rather than the final PRISMA-selected corpus.

The retained full-text evidence varies in research design, participant characteristics, disciplinary context, and outcome measures. It includes qualitative and process-oriented investigations, mixed-methods studies, quasi-experimental designs, and survey-based research. For example, Zhao et al. (2025) examined collaborative reflective discourse among pre-service science teachers, Sun and Huang (2026) investigated cognitive engagement during collaborative STEM lesson-plan refinement, and Stott and Scheepers (2026) examined engagement and perceived learning in a flipped-interaction intelligent-tutor approach (Stott & Scheepers, 2026; Sun & Huang, 2025; Zhao, 2025). These studies illustrate that the reviewed evidence addresses different learning processes and outcomes rather than a single, equivalent measure of learning effectiveness.

The populations also differ substantially. Directly relevant studies include pre-service science teachers engaged in collaborative reflection and science teaching activities science and mathematics student teachers participating in AI-supported teaching simulation and pre-service primary teachers undertaking AI-supported research-based learning (Ybyraimzhanov et al., 2026). Other studies examine pre-service mathematics teachers' lesson-planning reflections (Farillon-Labis, 2025). Ramnarain et al. (2025), by contrast, surveyed pre-service science teachers' intention to

use GenAI in inquiry-based teaching; this provides contextual evidence about readiness and adoption rather than evidence of an implemented learning intervention (Ramnarain et al., 2025).

The disciplinary contexts range from science and STEM lesson design, inquiry-oriented teaching, scientific teaching practice, and simulated teaching to digital storytelling, mathematics, language education, and general pedagogical development. This breadth helps identify potentially transferable mechanisms, but it also requires the synthesis to distinguish direct evidence from evidence that is only contextual. Results from non-science populations should not be generalised automatically to pre-service science teacher education (Ching Cheung et al., 2025).

The reviewed studies draw on different theoretical perspectives, including Kolb's experiential learning theory, ALACT, Schön's reflective practice, Learning by Design, the 5E model, transformative learning, self-determination theory, ICAP, the Theory of Planned Behavior, and metacognitive self-regulated learning. The presence of experience, reflection, or revision activities does not by itself demonstrate implementation of Kolb's cycle. Several studies use alternative frameworks or refer to Kolb only as a theoretical or interpretive lens.

Reflection-related evidence also varies. Some studies use direct reflective-thinking questionnaires, while others analyse reflective journals, written reflections, interviews, discourse patterns, or interaction logs. These forms of evidence represent different constructs and levels of inference. For example, Wei et al. (2025) used an adapted 41-item reflective-thinking instrument, and Pansamut et al. (2026) used a 12-item reflective-thinking questionnaire. By contrast, Zhao et al. (2025) examined collaborative reflective discourse, and Choi and Kim (2026) analysed journals, AI interaction records, lesson-design artefacts, and microteaching-related materials. Such qualitative evidence can illuminate reflective practice but should not be described as a measured change in reflective-thinking ability (Wei et al., 2025).

Overall, the evidence indicates an emerging but fragmented field. GenAI has been incorporated into dialogue, lesson planning, inquiry support, simulation, and reflective activities, but the reviewed studies differ in population, theoretical grounding, learning sequence, and outcome measurement. This heterogeneity requires the three research questions to be answered separately and requires careful boundaries around what can be claimed for science teacher education.

**Table 2.** Characteristics of the Reviewed Studies and Contextual Evidence

| Characteristic | Description from the current synthesis |
|---|---|
| Research designs | Qualitative case studies, phenomenological and process-oriented studies, mixed-methods research, quasi-experimental studies, comparative analyses, and survey-based research |
| Participant populations | Pre-service science teachers, science student teachers, elementary pre-service teachers, mathematics and language teacher candidates, educational-technology students, and candidates from social studies, physical education, and general teacher education |
| Disciplinary contexts | Science and STEM education, elementary science, mathematics, language education, educational technology, social studies, physical education, and general teacher education |
| GenAI applications | ChatGPT-supported reflection and lesson planning, GenAI-supported lesson redesign, research-based learning, interactive tutoring, virtual teaching simulation, digital storytelling, and metacognitive scaffolding |
| Theoretical frameworks | Kolb, ALACT, Schön, Learning by Design, 5E, transformative learning, self-determination theory, ICAP, Theory of Planned Behavior, and self-regulated learning |

| | |
|---|---|
| Reflection-related evidence | Reflective-thinking questionnaires, reflective discourse analysis, journals, interviews, learning logs, written reflections, engagement measures, and perceived-learning measures |

Note: The table summarises the currently assessed working evidence. It is not the final profile of the complete SLR corpus; ten priority records remain pending assessment.

### *Applications of generative AI in teacher education (RQ1)*

The reviewed studies indicate that GenAI has been used in several interconnected ways in teacher education. Its roles include initiating reflective dialogue, supporting instructional design, scaffolding cognitive and metacognitive activity, facilitating inquiry or research-based learning, and providing simulated or interactive teaching support. The role of the tool depends on how learners interact with it and on the pedagogical task in which it is embedded.

### *GenAI as a dialogic partner and reflective scaffold*

One recurring application of GenAI involves its use as a dialogic partner that supports reflection through questioning, alternative perspectives, feedback, and iterative discussion. In this configuration, the AI system does not simply provide information but participates in a reflective process in which learners reconsider their experiences, examine assumptions, and generate alternative instructional responses.

Zhao et al. (2025) provide a particularly relevant example. Their study involved 14 pre-service science teachers who participated in collective reflection following microteaching and science teaching activities. ChatGPT 4.0 was provided to two of the four reflection groups as a thought partner. The chatbot supported three ALACT-informed activities: looking back on teaching experiences, identifying essential aspects of teaching situations, and generating alternative methods of action. The ChatGPT-supported groups demonstrated more multivoiced discourse than groups without ChatGPT. However, they also showed lower proportions of pedagogical reasoning and supportive critique. Thus, a conversational AI partner may broaden the range of perspectives without necessarily ensuring deeper critical synthesis.

Abualrob (2025) examined ChatGPT-assisted re-reflection among pre-service elementary teachers during a science teaching practicum. Participants revisited traditional journal entries through interaction with ChatGPT and then produced revised reflections. The chatbot served as a reflective scaffold by prompting participants to reconsider assumptions and connect teaching experiences with alternative pedagogical interpretations. Erbay-Çetinkaya (2025), Arefian et al. (2026), also examined ChatGPT as a reflective partner in collaborative teacher-learning environments. These studies suggest that GenAI may support reflective dialogue, although the depth and quality of reflection still depend on learners' evaluation of AI suggestions and their connection to authentic teaching experiences (Arefian et al., 2025, 2026; Erbay-Çetinkaya, 2025).

Nevertheless, the findings also reveal limitations. AI-generated questions and interpretations may be generic, repetitive, or insufficiently sensitive to the particular classroom context. Consequently, the reflective value of chatbot interaction depends on participants' ability to evaluate AI suggestions and relate them to actual teaching experiences.

### *GenAI as an instructional design and lesson-planning assistant*

Another prominent application concerns the use of GenAI in preparing, evaluating, and refining instructional materials. In these studies, pre-service teachers interacted with AI systems to generate lesson ideas, formulate learning activities, identify possible student

misconceptions, organize instructional sequences, and improve the alignment between learning objectives and teaching strategies.

Sun and Huang (2026) examined guided ChatGPT-supported collaborative refinement of STEM lesson plans among 12 pre-service teachers. An embedded GPT-4 chatbot served as a shared resource for six student pairs. Participants examined existing lesson plans, discussed AI-generated suggestions, and revised their instructional designs. Their findings showed that the use of structured prompts was associated with different patterns of cognitive engagement, including analysis, reflection, and synthesis. However, the study treated reflection as one category within a broader cognitive-engagement framework rather than as a separate reflective-thinking outcome (Sun & Huang, 2026).

Gunbatar et al. (2025) investigated how participants with different levels of teaching experience interacted with ChatGPT when preparing science lessons on acids and bases. Their findings identified four interaction profiles: AI-reliant planners, AI-collaborative planners, AI-assisted plan refiners, and AI-independent planners. The two pre-service science teachers in the study tended to rely more heavily on ChatGPT suggestions, whereas experienced teachers demonstrated greater selectivity and critical evaluation. This finding highlights the influence of prior pedagogical and content knowledge on how teachers interpret and incorporate AI-generated suggestions (Gunbatar et al., 2025).

Other studies examined ChatGPT-assisted lesson planning and design among elementary and science pre-service teachers. Choi (2026) investigated motivational manifestations during ChatGPT-assisted elementary science lesson planning, while Katsampoxaki-Hodgetts (2026) examined how pre-service science teachers used GenAI tools to redesign lessons and justify instructional decisions (Choi, 2026; Katsampoxaki-Hodgetts, 2026b). Choi and Kim (2026) studied an eight-participant, ten-week ChatGPT-assisted science lesson-design programme for elementary pre-service teachers. The programme included iterative AI interaction, lesson-plan revision, peer discussion, microteaching, and reflective journals. Although these activities created opportunities for reflection and revision, the study did not use a direct reflective-thinking scale or explicitly implement Kolb's four-stage cycle (Choi & Kim, 2026).

The use of multiple GenAI platforms was also documented in mathematics teacher education. Farillon-Labis (2025) examined pre-service mathematics teachers' reflections while using ChatGPT, Gemini, and Perplexity AI for lesson planning. Participants compared AI-generated responses, refined prompts, and exercised pedagogical judgment when deciding whether to adopt or reject suggestions. Although the subject context differs from science education, the findings provide supporting evidence regarding the role of AI literacy and evaluative judgment in instructional design.

***GenAI as a cognitive and metacognitive scaffold***

A further application involves the use of GenAI to support cognitive organization, metacognitive monitoring, and self-regulated learning. In this role, AI systems provide prompts, feedback, suggestions, or structured guidance intended to help learners organize their reasoning and monitor their progress during complex learning tasks.

Pansamut et al. (2026) investigated an AI-integrated design–reflection–revision process involving 60 pre-service teachers from social studies, physical education, and dance education(Pansamut et al., 2026). A 12-item reflective-thinking questionnaire indicated an increase in participants' reported reflective-thinking scores following the intervention. However, the participant population was not pre-service science teachers,

and the learning sequence was not a complete implementation of Kolb's four stages. The study therefore provides supporting evidence about reflective-thinking measurement in GenAI-supported teacher education rather than direct evidence for science teacher preparation.

Dahri et al. (2024) examined ChatGPT acceptance and its relationship with metacognitive self-regulated learning among 300 pre-service teachers(Dahri et al., 2024). ChatGPT was used in scenario-based lesson-planning activities, while metacognitive self-regulated learning was examined through questionnaire data and qualitative reflections. The findings suggest that ChatGPT may support planning, monitoring, and post-lesson evaluation. However, metacognitive self-regulation should not be treated as equivalent to directly measured reflective-thinking ability.

Ybyraimzhanov et al. (2026) examined Gemini 2.5 Pro as a cognitive and metacognitive scaffold in a 15-week research-based learning course involving 280 pre-service primary teachers. AI support was integrated into five stages: problem framing, hypothesis formulation, methodological design, literature review, and mock-data interpretation. The study reported improvements in creative scientific competence and analysed reflection journals from a subsample. These findings support the potential of AI-assisted inquiry and research learning, but creative scientific competence is not equivalent to directly measured reflective-thinking ability, and the five-stage sequence is not Kolb's cycle. Ramnarain et al. (2025), by contrast, examined pre-service science teachers' intention to use GenAI in inquiry-based teaching. Their survey and focus-group findings are relevant to adoption readiness and implementation conditions, but the study did not test a chatbot-supported learning intervention.

***GenAI as a simulation and creative production partner***

GenAI has also been integrated into simulation-based and interactive learning environments. Zhang et al. (2026) investigated a GenAI-powered virtual-reality simulation in which pre-service science and mathematics teachers practised dialogic teaching with virtual students. The activity enabled simulated teaching, interpretation of student responses, and reconsideration of pedagogical decisions, but the instructional design was framed primarily through dialogic teaching and the 5E model rather than a complete Kolb cycle. Stott and Scheepers (2026) studied 42 science student teachers using free GenAI platforms, mostly ChatGPT, configured as flipped-interaction intelligent tutors over ten weeks. The tutors prompted students with questions and hints and were compared with interactive worksheets. The study assessed engagement and perceived science learning, not reflective-thinking development.

Wei et al. (2025) examined GenAI-assisted digital storytelling among 80 pre-service teachers from educational-technology-related majors. ChatGPT supported story ideation and script development, while other generative tools were used to create visual and animated elements. The study used an adapted 41-item instrument to examine reflective-thinking skills and reported higher reflection frequencies across single-loop, double-loop, and triple-loop reflection in the experimental group. This provides direct measurement evidence, but the participants were not science/STEM teacher candidates and the storytelling sequence was not explicitly mapped to Kolb's cycle.

Taken together, these studies show that GenAI may function as a dialogue partner, instructional-design assistant, cognitive scaffold, inquiry-learning support, interactive tutor, simulation agent, or creative-production partner. The functions are not interchangeable, and their educational significance depends on the extent to which

participants evaluate AI outputs, connect them to disciplinary evidence, and retain responsibility for pedagogical decisions.

**Tabel 3.** Functions of GenAI in the reviewed teacher-education studies

| GenAI application | Main function | Representative studies | Evidence boundary |
|---|---|---|---|
| Dialogic reflective partner | Prompts reflection, offers alternative perspectives, and supports reflective dialogue | Zhao et al. (2025); Abualrob (2025); Erbay-Çetinkaya (2025); Arefian et al. (2026); Esfandiari and Arefian (2025); Choi and Kim (2026) | Reflection is often examined qualitatively; not all studies measure reflective thinking directly |
| Instructional design assistant | Generates lesson ideas, suggests activities, and supports lesson-plan evaluation and refinement | Sun and Huang (2026); Gunbatar et al. (2025); Choi (2026); Choi and Kim (2026); Katsampoxaki-Hodgetts (2026); Farillon-Labis (2025) | Lesson-design performance and reflective thinking are distinct outcomes |
| Cognitive and metacognitive scaffold | Provides feedback, prompts, and support for planning, monitoring, and revision | Pansamut et al. (2026); Dahri et al. (2024); Özbilen et al. (2026); Ybyraimzhanov et al. (2026) | Metacognition and self-regulated learning are related to, but not interchangeable with, reflective thinking |
| Simulation agent | Provides virtual teaching interactions and simulated student responses | Zhang et al. (2026) | Simulation-based practice does not automatically constitute a full Kolb cycle |
| Creative production partner | Supports idea generation, script development, and multimodal content creation | Wei et al. (2025) | Direct reflective-thinking measurement is present, but the subject context is not science/STEM |
| Collaborative learning resource | Supports group discussion, collaborative lesson design, and evaluation of learning products | Lehtinen et al. (2026); Li et al. (2025) | Evidence often concerns critical thinking or collaborative processes rather than direct reflective-thinking outcomes |
| Flipped-interaction tutor | Uses prompts, questions, and hints to support interactive science learning | Stott and Scheepers (2026) | Engagement and perceived learning were assessed; reflective thinking and Kolb's cycle were not directly measured |

| GenAI application | Main function | Representative studies | Evidence boundary |
|---|---|---|---|
| Research-based learning support | Scaffolds problem framing, hypothesis formation, methodological design, literature review, and data interpretation | Ybyraimzhanov et al. (2026) | Creative scientific competence and journal reflections are not equivalent to a direct reflective-thinking measure |
| Adoption and readiness evidence | Examines intention, perceived usefulness, AI literacy, and readiness to use GenAI in inquiry teaching | Ramnarain et al. (2025) | Contextual evidence about uptake; not an implemented chatbot-supported learning intervention |

Note: The categories are analytical groupings developed for this review. Individual studies may involve more than one GenAI function.

### *Alignment with experiential learning theory*

The reviewed studies demonstrate that GenAI has been incorporated into practice-based, inquiry-oriented, and reflective learning activities. However, the presence of experience, reflection, or revision does not necessarily indicate that an intervention follows Kolb's complete experiential learning cycle. The synthesis therefore distinguishes explicit implementation of the four stages from theoretical references to Kolb and from activities that show only partial conceptual overlap.

Zhao et al. (2025) used Korthagen's ALACT framework to structure collective reflection after teaching activities. The study included reflection on experience and the generation of alternative actions, but it did not implement Kolb's four stages. Abualrob (2025) referred to Kolb as an interpretive framework for understanding practicum-based re-reflection; the study did not prescribe a four-stage Kolb sequence. Zhang et al. (2026) discussed experiential learning in relation to virtual teaching simulation, but the intervention was primarily organised around dialogic teaching and the 5E model. Pansamut et al. (2026) used an AI-supported design–reflection–revision process and referred to Kolb conceptually, yet the four stages were not operationalised as a complete intervention.

Other studies also offered practice-based opportunities without explicit Kolb implementation. Ybyraimzhanov et al. (2026) organised AI-supported research-based learning into five stages, while Choi and Kim (2026) incorporated lesson design, iterative AI interaction, peer discussion, microteaching, and reflection. Sun and Huang (2026) focused on collaborative refinement of existing lesson plans rather than classroom enactment. Gunbatar et al. (2025) examined AI-supported science lesson planning, and Stott and Scheepers (2026) used flipped-interaction intelligent tutors to support science learning. These activities may contain elements resembling experience, reflection, conceptualisation, or experimentation, but the studies did not explicitly structure the full four-stage Kolb cycle.

Overall, the current evidence indicates that experiential and reflective elements are present in GenAI-supported teacher education, but explicit integration of all four Kolb stages remains limited in the studies assessed. Therefore, these studies should not be described collectively as Kolb-based interventions. The more defensible conclusion is that GenAI-supported teacher learning has begun to incorporate practice and reflection, while a systematic cycle linking experience, reflective observation, conceptual

understanding, and active experimentation remains an important design and research gap.

**Table 4**. Relationship between GenAI-supported learning activities and Kolb's experiential learning c ycle

| Study | Reported learning activity or framework | Relationship with Kolb |
|---|---|---|
| Zhao et al. (2025) | Microteaching and science teaching followed by collective reflection; ALACT | Alternative reflective framework; not a four-stage Kolb intervention |
| Abualrob (2025) | Science teaching practicum, journal re-reflection, conceptual reframing, and pedagogical alternatives | Kolb used as an interpretive lens; stages not prescribed as a complete sequence |
| Zhang et al. (2026) | Virtual-reality dialogic teaching simulation; 5E-oriented activity | Practice-based simulation; not explicitly mapped to all four Kolb stages |
| Pansamut et al. (2026) | AI-supported design–reflection–revision | Partial conceptual overlap; Kolb cited but not operationalised as a full cycle |
| Ybyraimzhanov et al. (2026) | Five-stage AI-supported research-based learning | Inquiry sequence, but not Kolb's four-stage cycle |
| Choi and Kim (2026) | Science lesson design, iterative AI dialogue, peer discussion, microteaching, and reflection | Practice and reflection elements; no explicit complete Kolb cycle |
| Sun and Huang (2026) | Collaborative refinement of existing STEM lesson plans | Lesson refinement without teaching enactment; no explicit Kolb cycle |
| Gunbatar et al. (2025) | ChatGPT-supported science lesson planning | Planning activity; no classroom enactment or explicit Kolb sequence |
| Stott and Scheepers (2026) | Flipped-interaction intelligent tutoring and interactive worksheets | Interactive learning support; no explicit Kolb cycle |

Note: "Partial conceptual overlap" refers to activities that resemble one or more elements of experiential learning; it does not mean that the study implemented the full cycle.

### *Reflective-thinking evidence, assessment, implementation challenges, and SDG 4*

#### *Evidence and assessment of reflective thinking*

The evidence for reflective-thinking development is uneven. A small number of studies used instruments explicitly designed to measure reflective thinking, while many science-related studies examined reflective practice through journals, interviews, interaction logs, or discourse analysis. These qualitative sources provide useful evidence about how participants reflect, but they do not by themselves demonstrate a quantified change in reflective-thinking ability.

Wei et al. (2025) used an adapted 41-item reflective-thinking instrument in a study of GenAI-assisted digital storytelling among pre-service teachers in educational-technology-related majors. Pansamut et al. (2026) used a 12-item reflective-thinking questionnaire in an AI-supported design–reflection–revision activity with pre-service teachers from social studies, physical education, and dance education. Both studies offer direct questionnaire-based evidence, but neither focuses on pre-service science teachers. Their findings should therefore be treated as supporting evidence rather than direct confirmation of reflective-thinking development in the target population.

In science-related teacher education, Zhao et al. (2025) analysed collaborative reflective discourse using network-based and reflection-level coding. Their results

showed more multivoiced discourse in the ChatGPT-supported groups, alongside lower proportions of pedagogical reasoning and supportive critique than in the non-AI groups. Abualrob (2025) examined changes in the quality of written re-reflection during a science practicum, while Choi and Kim (2026) analysed reflection journals, interaction logs, lesson plans, and microteaching-related evidence. Ybyraimzhanov et al. (2026) included reflection journals from a subsample, but the principal quantitative outcome was creative scientific competence, not reflective-thinking ability. These studies provide process evidence about reflection and learning, not equivalent scale-based measurements of reflective-thinking development.

**Table 5.** Types of reflection-related evidence and their interpretive boundaries

| Study | Evidence or instrument | What the evidence supports | Interpretive boundary |
|---|---|---|---|
| Wei et al. (2025) | Adapted 41-item reflective-thinking instrument | Direct questionnaire-based reflective-thinking evidence | Educational-technology-related majors; not science/STEM teacher preparation |
| Pansamut et al. (2026) | 12-item reflective-thinking questionnaire; learning logs | Direct questionnaire-based evidence of reported reflective thinking | Participants from social studies, physical education, and dance education; not science specialists |
| Zhao et al. (2025) | Epistemic Network Analysis and reflection-level discourse coding | Evidence about reflective discourse and interaction patterns | Not a standardised reflective-thinking ability scale |
| Abualrob (2025) | Traditional and revised journals; Onion Model analysis | Qualitative evidence of changes in reflection quality | Practicum reflection among elementary pre-service teachers; no standardised scale |
| Choi and Kim (2026) | Interaction logs, reflective journals, lesson plans, observations, and reports | Qualitative evidence of AI literacy, evaluation, and reflective practice | Small voluntary programme; no direct reflective-thinking scale |
| Ybyraimzhanov et al. (2026) | Creative Scientific Competence Test and reflection journals | Evidence of creative scientific competence and reported reflection processes | Creative competence is not reflective-thinking ability |
| Stott and Scheepers (2026) | Engagement items, perceived science-learning ratings, essays, and interviews | Evidence about engagement and perceived learning | Reflective thinking was not directly measured |

### *Implementation challenges*

The reviewed studies identify several conditions that may constrain the educational value of GenAI. First, AI-generated responses may contain inaccuracies or hallucinations, requiring learners to verify claims against disciplinary sources. Gunbatar et al. (2025) reported differences in how participants with different levels of teaching experience evaluated AI suggestions; the pre-service teachers tended to rely

more heavily on ChatGPT, whereas more experienced teachers were more selective. Stott and Scheepers (2026) also reported user-identified errors and concerns about workload and repetitive interactions in the flipped-tutor approach.

Second, the presence of AI dialogue does not guarantee deeper reasoning. Zhao et al. (2025) found more multivoiced discourse in AI-supported groups but lower proportions of pedagogical reasoning and supportive critique. Third, small samples, voluntary participation, simulated or planned teaching rather than authentic classroom enactment, and reliance on self-reported outcomes limit the strength and transferability of some findings. Finally, studies of intention and readiness, such as Ramnarain et al. (2025), indicate that AI literacy, perceived usefulness, skills, and concerns need to be considered before GenAI-supported inquiry is implemented at scale.

***Implications for quality science teacher education and SDG 4***

The findings of this review provide insights into the potential contribution of GenAI-supported learning to the broader goal of quality education under Sustainable Development Goal 4 (SDG 4). The reviewed studies demonstrate that GenAI can support several aspects of pre-service teacher education, including collaborative reflection, instructional design, inquiry-based learning, pedagogical decision-making, and the development of digital and AI literacy. These applications are relevant to efforts to prepare future science teachers who can critically evaluate AI-generated information and integrate digital technologies into meaningful learning experiences.

However, the findings also indicate that the use of GenAI does not automatically ensure the development of reflective thinking or improve the quality of teacher preparation. The effectiveness of GenAI-supported learning depends on how AI interactions are integrated into pedagogically meaningful activities, how learners evaluate AI-generated responses, and whether opportunities for experience, reflection, conceptual understanding, and further experimentation are systematically provided. Therefore, the potential contribution of GenAI to SDG 4 lies not merely in the availability of AI tools but in their responsible and purposeful integration into teacher education.

From this perspective, the review highlights the need for learning designs that combine GenAI support, experiential learning, and explicit reflective-thinking activities in science teacher education. Such designs may contribute to more meaningful, inclusive, and digitally responsive teacher preparation. Nevertheless, the reviewed studies did not directly measure the achievement of SDG 4 outcomes. Their contributions should therefore be interpreted as educational implications and directions for future research rather than demonstrated impacts on SDG 4.

For science teacher education, a future GenAI-supported experiential design could connect the four Kolb stages more explicitly. Concrete experience could involve a science investigation, field inquiry, engineering-design task, or microteaching episode. During reflective observation, the chatbot could ask learners to compare predictions with observations, identify inconsistencies, and justify interpretations. During abstract conceptualisation, it could help learners connect evidence to scientific concepts while requiring verification against reliable sources. During active experimentation, learners could redesign an investigation, prototype, or teaching strategy and test the revised approach. This proposed sequence preserves learner responsibility and should be examined empirically; it is not a pattern already established across the reviewed studies.

## *Discussion*

### *GenAI as a learning resource rather than a substitute for professional judgement*

The findings for RQ1 show that GenAI is being used for dialogue, lesson planning, reflective scaffolding, inquiry support, simulation, and creative production. These functions create different opportunities for teacher learning. In particular, dialogue and iterative feedback may help pre-service teachers consider alternatives and articulate pedagogical decisions. However, AI assistance should not be equated with improved professional reasoning. The quality of learning depends on the authenticity of the task, the structure of the prompts, the participants' prior knowledge, and their capacity to question and verify generated content.

### *The gap between experiential activities and a structured Kolb cycle*

The RQ2 synthesis highlights a distinction between practice-based learning and theoretically structured experiential learning. Microteaching, practicum experience, research-based learning, virtual simulation, lesson design, and revision may provide meaningful learning experiences. Yet these activities do not become Kolb-based merely because learners experience an activity and subsequently reflect on it. A more rigorous design should intentionally connect concrete experience, reflective observation, abstract conceptualisation, and active experimentation, with clear tasks and evidence at each stage. The limited explicit mapping found in the reviewed evidence represents a gap for future science teacher education research.

Reflective-thinking development requires construct-aligned assessment

The RQ3 findings indicate that direct measurement of reflective thinking remains limited in the science/STEM teacher-education evidence. Studies that use reflective journals, discourse analysis, interaction logs, or self-reported engagement provide valuable insight into reflective practice and learning processes, but they should not be interpreted as interchangeable with validated measures of reflective-thinking ability. Future studies should combine process data with construct-aligned instruments, performance tasks, and transparent scoring procedures. They should also distinguish reflective thinking from metacognition, self-regulated learning, critical thinking, AI literacy, and creative competence.

### *SDG 4 and responsible integration*

The SDG 4 lens draws attention to the conditions under which GenAI could contribute to inclusive and high-quality teacher preparation. The findings of this review suggest that GenAI-supported activities may provide opportunities to strengthen pre-service teachers' pedagogical competence, digital and AI literacy, reflective engagement, and capacity to make evidence-informed instructional decisions. These potential contributions, however, depend on intentional curriculum design, equitable access to digital technologies, critical digital literacy, ethical use, and appropriate human oversight. The findings also highlight the importance of integrating GenAI into meaningful science learning activities rather than treating the availability of AI tools as an indicator of educational quality. Because the reviewed studies rarely measured outcomes directly related to SDG 4, the implications should be interpreted as potential educational contributions and directions for future research, rather than evidence that GenAI use itself advances SDG attainment.

## CONCLUSION

**Fundamental Finding:** This systematic literature review demonstrates the growing potential of Generative Artificial Intelligence (GenAI) to transform learning experiences in pre-service teacher education through diverse pedagogical and cognitive functions. The reviewed studies indicate that GenAI serves not only as an information-retrieval tool but also as a dialogic learning partner, instructional-design assistant, cognitive and metacognitive scaffold, inquiry-learning facilitator, interactive tutor, simulation agent, and creative-production resource. These multifaceted roles highlight its capacity to facilitate knowledge construction, pedagogical decision-making, and reflective engagement. Nevertheless, the educational value of GenAI is determined not simply by its adoption but by the extent to which it is purposefully integrated into pedagogically meaningful learning experiences. The findings further reveal that the systematic integration of GenAI with all four stages of Kolb's experiential learning cycle remains underexplored, particularly in pre-service science and STEM teacher education. **Implication:** The findings underscore the importance of designing GenAI-supported learning environments that establish a meaningful connection between experiential learning, reflective thinking, and professional teacher development. In particular, integrating GenAI into concrete experience, reflective observation, abstract conceptualization, and active experimentation may provide opportunities for pre-service teachers to develop critical reflection, pedagogical reasoning, and informed instructional decision-making. This integration is particularly relevant to science teacher education, where inquiry, experimentation, and evidence-based reasoning constitute essential dimensions of professional practice. Furthermore, from the perspective of Sustainable Development Goal 4 (SDG 4), the findings emphasize the potential contribution of responsible GenAI integration to preparing future teachers who can promote inclusive, equitable, and quality education. However, this contribution should be interpreted as a prospective educational implication rather than a directly measured outcome of the reviewed studies. **Limitation:** Several limitations should be acknowledged. First, the reviewed studies exhibit considerable heterogeneity in educational settings, participant characteristics, GenAI applications, pedagogical approaches, and research methodologies, which limits direct comparisons and the synthesis of findings across contexts. Second, the relatively limited representation of pre-service science and STEM teacher education restricts the generalizability of the findings to these specific populations. Third, reflective practice has predominantly been examined through qualitative approaches, with insufficient application of validated instruments specifically designed to assess reflective-thinking development. Consequently, the extent to which GenAI-supported learning directly contributes to improvements in reflective thinking remains difficult to establish. These limitations indicate that the findings should be interpreted within the methodological and contextual boundaries of the available evidence. **Future Research:** Future investigations should develop and empirically evaluate integrated pedagogical frameworks that combine GenAI-supported learning, the complete four-stage Kolb experiential learning cycle, and explicit assessment of reflective thinking among pre-service science and STEM teachers. Such frameworks should be implemented in authentic educational contexts, including science inquiry, instructional planning, microteaching, and teaching practicum, to examine how GenAI facilitates reflective processes, pedagogical reasoning, and professional development. Further research should employ validated, construct-aligned assessment instruments and adopt longitudinal or mixed-methods designs to provide stronger evidence of learning outcomes and developmental changes.

Additionally, investigations should address potential challenges associated with AI-generated inaccuracies, algorithmic limitations, academic integrity, and excessive learner dependence on AI. Ultimately, future research should move beyond examining the extent of GenAI adoption toward establishing how its pedagogically grounded integration can cultivate reflective, critical, and responsible future teachers capable of harnessing emerging technologies to advance the broader goals of quality education.

## AUTHOR CONTRIBUTIONS

**An Nuril Maulida Fauziah** (Corresponding Author) contributed to conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, and project administration. **Tutut Nurita** contributed to conceptualization, methodology, validation, supervision, and writing – review and editing. **Ahmad Fauzi Hendratmoko** contributed to investigation, data curation, formal analysis, visualization, and writing – review and editing. **Martini** contributed to investigation, validation, and writing – review and editing. **Enny Susiyawati** contributed to methodology, validation, and writing – review and editing. **Dyah Astriani** contributed to validation, supervision, and writing – review and editing. **Erman** contributed to conceptualization, supervision, and writing – review and editing. **Fatimatuz Zahro** contributed to investigation, data curation, formal analysis, visualization, and writing – original draft.
All authors participated in the preparation of this manuscript, critically reviewed the intellectual content, and approved the final version for submission.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no financial or personal conflicts of interest that could have influenced the research, analysis, interpretation, or publication of this study.

## ETHICAL COMPLIANCE STATEMENT

This study is a systematic literature review based on previously published studies and publicly available academic literature. It did not involve direct participation of human subjects, animal experiments, or the collection of identifiable personal data. Therefore, ethical approval was not required. The authors confirm that this manuscript is original, has been prepared in accordance with established research and publication ethics, and appropriately acknowledges all sources used. All authors take responsibility for the integrity and accuracy of the work.

## STATEMENT ON THE USE OF AI OR DIGITAL TOOLS IN WRITING

The authors acknowledge the use of artificial intelligence (AI)-based and digital tools to support the preparation and writing of this manuscript. Specifically, ChatGPT (OpenAI) was used to assist with language refinement, academic writing improvement, manuscript structuring, and the clarification of research arguments. These tools were used as writing-support instruments and not as substitutes for the authors' scholarly judgment, critical evaluation, or research decisions. All AI-assisted outputs were critically reviewed, verified, and revised by the authors to ensure accuracy, academic integrity, and consistency with the reviewed literature. The authors take full responsibility for the content, interpretation, and final version of this manuscript.

## REFERENCES


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

Arefian, M. H., Çomoğlu, I., & Dikilitaş, K. (2026). Understanding EFL teachers' experiences of ChatGPT-driven collaborative reflective practice through a community of practice lens. *Innovation in Language Learning and Teaching*, *20*(2), 318–333. https://doi.org/10.1080/17501229.2024.2412769

Arefian, M. H., Esfandiari, R., & Zarei, A. A. (2025). Exploring AI-based collaborative reflective practice in light of ChatGPT: Insights from EFL preservice teachers. *Applied Linguistics Inquiry*, *3*(1), 31–43. https://doi.org/10.22077/ali.2025.9022.1108

Beale, R. (2025). *Dialogic pedagogy for large language models: Aligning conversational AI with proven theories of learning* (arXiv:2506.19484). arXiv. https://doi.org/10.48550/arXiv.2506.19484

Breslyn, W., & McGinnis, J. R. (2019). Investigating preservice elementary science teachers' understanding of climate change from a computational thinking systems perspective. *Eurasia Journal of Mathematics, Science and Technology Education*, *15*(6), Article em1696. https://doi.org/10.29333/ejmste/103566

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, *368*, Article l6890. https://doi.org/10.1136/bmj.l6890

Chang, D. H., Lin, M. P. C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, *15*(17), Article 12921. https://doi.org/10.3390/su151712921

Cheung, K. K. C., Zerouali, A., Koenen, J., & Erduran, S. (2025). Do generative artificial intelligence (GenAI) and science education mix? A systematic review of the literature. *Studies in Science Education*. Advance online publication. https://doi.org/10.1080/03057267.2025.2578091

Choi, Y.-S. (2026). Exploring pre-service elementary teachers' motivational manifestations in ChatGPT-assisted lesson planning: A self-determination theory perspective. *Asia-Pacific Science Education*. Advance online publication. https://doi.org/10.1163/23641177-bja10108

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

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A., & Soomro, R. B. (2024). Extended TAM based acceptance of AI-powered ChatGPT for supporting metacognitive self-regulated learning in education: A mixed-methods study. *Heliyon*, *10*(8), Article e29317. https://doi.org/10.1016/j.heliyon.2024.e29317

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, *20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

Erbay-Çetinkaya, Ş. (2025). A triadic support framework for reflective teacher identity: Peer, AI, and faculty collaboration within a community of practice. *Teaching and Teacher Education*, *168*, Article 105239. https://doi.org/10.1016/j.tate.2025.105239

Farillon-Labis, S. J. (2025). Exploring pre-service mathematics teachers' reflections in lesson planning with generative artificial intelligence. *International Journal of Learning, Teaching and Educational Research*, *24*(11), 334–355. https://doi.org/10.26803/ijlter.24.11.16

Gunbatar, S. A., Sirin, G. T., Ilkyaz, O. C., & Mutlu, Y. (2025). Exploring the artificial intelligence interaction profiles of participants with different levels of teaching experience for lesson planning in the context of acids and bases. *Chemistry Education Research and Practice*, *26*(4), 977–995. https://doi.org/10.1039/D5RP00064E

Hordvik, M., Fletcher, T., Haugen, A. L., Møller, L., & Engebretsen, B. (2024). The process(es) of learning about teaching using models-based practice: Pre-service teachers' experiences. *European Physical Education Review*, *31*(3), 444–461. https://doi.org/10.1177/1356336X241282492

Johnston, R. B. (2016). Arsenic and the 2030 Agenda for Sustainable Development. In P. Bhattacharya, M. Vahter, J. Jarsjö, J. Kumpiene, A. Ahmad, C. Sparrenbom, G. Jacks, M. E. Donselaar, J. Bundschuh, & R. Naidu (Eds.), *Arsenic research and global sustainability: Proceedings of the Sixth International Congress on Arsenic in the Environment (As2016)* (pp. 12–14). CRC Press. https://doi.org/10.1201/b20466-7

Katsampoxaki-Hodgetts, K. (2026). Enacted AI literacy towards epistemic visibility: Embedding human-AI design as assessment-as-learning in pre-service science teachers' lesson redesign. *Research in Science Education*. Advance online publication. https://doi.org/10.1007/s11165-026-10360-7

Kolb, D. A. (2015). *Experiential learning: Experience as the source of learning and development* (2nd ed.). Pearson Education.

Krushinskaia, K., Elen, J., & Raes, A. (2026). Pre-service teachers' agency during their interactions with generative AI while designing for learning: A process view on Intelligent-TPACK. *Computers and Education Open*, *10*, Article 100325. https://doi.org/10.1016/j.caeo.2025.100325

Lehtinen, A., Channa, F., Näykki, P., Ahlström, E., & Hiljanen, M. (2026). Generative AI for collaborative learning: Fostering critical thinking in teacher education. *Journal of Computer Assisted Learning*, *42*(3), Article e70256. https://doi.org/10.1002/jcal.70256

Li, H., Chai, C. S., & Wang, X. (2025). Promoting pre-service music teachers' TPACK with generative AI: An intervention through designing with AI. *Computers and Education: Artificial Intelligence*, *9*, Article 100525. https://doi.org/10.1016/j.caeai.2025.100525

Özbilen, F. M., Asghar, M. Z., & Bayrakci, M. (2026). A multimethod exploration of human-AI-supported shared metacognition for inclusive educational leadership during crisis in Türkiye and Pakistan. *Discover Computing*, *29*, Article 354. https://doi.org/10.1007/s10791-026-10235-5

Pansamut, O. T., Phayakka, P., & Wang, X. (2026). AI-integrated reflective learning design and pre-service teachers' competency development: A mixed-methods study. *Malaysian Journal of Learning and Instruction*, *23*(2), 42–61. https://doi.org/10.32890/mjli2026.23.2.3

Ramnarain, U., Ogegbo, A. A., Penn, M., Ojetunde, S., & Mdlalose, N. (2025). Pre-service science teachers' intention to use generative artificial intelligence in

inquiry-based teaching. *Journal of Science Education and Technology*, *34*(6), 1272–1285. https://doi.org/10.1007/s10956-024-10159-z

Sherwood, G. (2024). Reflective practice and knowledge development: Transforming research for a practice-based discipline. *International Journal of Nursing Sciences*, *11*(4), 399–404. https://doi.org/10.1016/j.ijnss.2024.08.002

Spasopoulos, T., Sotiropoulos, D., & Kalogiannakis, M. (2025). Generative AI in pre-service science teacher education: A systematic review. *Advances in Mobile Learning Educational Research*, *5*(2), 1501–1523. https://doi.org/10.25082/amler.2025.02.007

Stott, A. E., & Scheepers, S. J. (2026). Turning free-to-use generative AI into flipped-interaction intelligent tutors: Exploring student engagement and perceptions of learning. *Canadian Journal of Science, Mathematics and Technology Education*, *26*(4), Article 100. https://doi.org/10.1007/s42330-026-00523-z

Sun, Y., & Huang, Y. (2025). Exploring the role of generative AI in collaborative lesson planning for pre-service teachers. *Journal of Computer Assisted Learning*, *41*(4), Article e70098. https://doi.org/10.1111/jcal.70098

Sun, Y., & Huang, Y. (2026). Exploring cognitive engagement in guided ChatGPT-supported collaborative lesson plan refinement for preservice teachers. *Educational Technology Research and Development*. Advance online publication. https://doi.org/10.1007/s11423-026-10657-4

Wei, X., Wang, L., Koszalka, T. A., Lee, L.-K., & Liu, R. (2025). Enhancing pre-service teachers' reflective thinking skills through generative AI-assisted digital storytelling creation: A three-dimensional framework analysis. *Computers & Education*, *235*, Article 105356. https://doi.org/10.1016/j.compedu.2025.105356

Ybyraimzhanov, K., Shakarim, N., Taurbekova, A., Kystaubayeva, B., Yessimbekova, A., & Kriviradeva, B. (2026). Enhancing pre-service teachers' creative scientific competence through AI-supported research-based learning. *Jurnal Pendidikan IPA Indonesia*, *15*(2), 197–209. https://doi.org/10.15294/jpii.v15i2.36717

Zhang, N., Ke, F., Dai, C. P., Southerland, S. A., & Barrett, A. (2026). Science and mathematics preservice teachers' perceptions and experiences of practicing dialogic teaching in generative AI-powered virtual reality simulation. *Teaching and Teacher Education*, *171*, Article 105349. https://doi.org/10.1016/j.tate.2025.105349

Zhao, J. (2025). Advancing English language education: A mixed-methods analysis of AI-driven tools' impact on engagement, personalization, and learning outcomes. *Education and Information Technologies*, *30*(15), 21773–21813. https://doi.org/10.1007/s10639-025-13560-2

*** Dr. An Nuril Maulida Fauziah, S.Pd., M.Pd. (Corresponding Author)**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: annurilfauziah@unesa.ac.id

**Dr. Tutut Nurita, S.Pd., M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: tututnurita@unesa.ac.id

**Dr. Ahmad Fauzi Hendratmoko, M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: ahmadhendratmoko@unesa.ac.id

**Dra. Martini, M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: martini@unesa.ac.id

**Enny Susiyawati, S.Si., M.Sc., M.Pd., Ph.D.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: ennysusiyawati@unesa.ac.id

**Dr. Dyah Astriani, S.Pd., M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: dyahastriani@unesa.ac.id

**Prof. Dr. Erman, M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: erman@unesa.ac.id

**Fatimatuz Zahro, S.Pd., M.Pd.**
Department of Science Education, Faculty of Mathematics and Natural Sciences, State University of Surabaya
Jl. Ketintang Selatan, Surabaya, East Java, Indonesia
Email: fatimatuz.20078@mhs.unesa.ac.id