MALAYSIAN JOURNAL OF LEARNING AND INSTRUCTION
https://e-journal.uum.edu.my/index.php/mjli

How to cite this article:
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

AI-INTEGRATED REFLECTIVE LEARNING DESIGN AND PRE-SERVICE TEACHERS' COMPETENCY DEVELOPMENT: A MIXED-METHODS STUDY

[1]Onnipa Thaithae Pansamut, [2]Phatchara Phayakka & [3]Xiaozhuan Wang
[1]School of Education & Research Center of Geosocial and Cultural Research for Sustainable Development, Walailak University, Nakhon Si Thammarat, Thailand
[2]School of Education, Walailak University, Nakhon Si Thammarat, Thailand
[3]School of Humanities, Ningbo University of Finance and Economics, Zhejiang, China

[2]Corresponding author: Phatchara.ph@mail.wu.ac.th

Received: 22/2/2026 Revised: 12/6/2026 Accepted: 7/7/2026 Published: 31/7/2026

ABSTRACT

Purpose – This study examined the developmental change in pre-service teachers' learning design competency and reflective thinking associated with their participation in an AI-integrated reflective learning design, and explored the reflective processes through which generative AI was used during instructional design. The study did not claim that the design would result in improvement, but it did provide a preliminary, within-group evidence and a process-oriented account.

Methodology – A convergent mixed-methods design was employed with 60 pre-service teachers enrolled in a learning design course. The structured learning design embedded generative AI within iterative cycles of design, reflection, and revision. Quantitative data were collected using a rubric-based learning design competency assessment and a 12-item reflective thinking questionnaire administered before and after participation, which was subsequently analysed using paired-samples t-tests. Qualitative data from 60 learning logs and 60 AI interaction logs (284 coded excerpts) were analysed using reflexive thematic analysis (Braun & Clarke, 2006, 2021). The two strands were merged through a joint display.

Findings – Scores on overall learning design competency and reflective thinking showed a significant increase from pretest to post-test, with the largest change in reflection-on-action. Qualitative findings

indicated that participants were engaged with generative AI as a reflective co-thinking partner through cycles of evaluation, modification, and rejection of AI-generated suggestions. However, there were also admissions of confusion, over-reliance or resistance among a small number of respondents.

Significance – The study contributes to a proposed reflective-mediation account that conceptualises generative AI as a pedagogically mediated tool rather than an efficiency-driven technology. For practice, the findings inform curriculum designers in teacher education by showing how reflective scaffolds (structured learning and AI interaction logs) can be embedded to support professional judgment. Given the single-group design, the contribution is exploratory and practice-oriented and warrants confirmation through comparative research.

Keywords: Generative artificial intelligence, pre-service teachers, learning design competency, reflective mediation, reflective scaffolding, AI as a co-thinking partner.

INTRODUCTION

The rapid emergence of generative artificial intelligence (AI), particularly large language models such as ChatGPT, has introduced new possibilities as well as concerns within teacher education. AI tools are increasingly positioned to support lesson planning, feedback generation, and instructional efficiency (van den Berg & du Plessis, 2023; Cai et al., 2025). For teacher education, however, the pressing question is not whether pre-service teachers are able to utilise or operate AI tools, but how such tools shape the formation of pedagogically grounded instructional reasoning.

This question reflects a concrete practical problem. Recent studies indicate that pre-service teachers often approach generative AI with ambivalence, recognising its capacity to generate instructional ideas while also expressing concerns about reliability, bias, contextual misalignment, and ethical implications (Bae et al., 2024; Espino, 2025; Liu et al., 2025). Many teacher education programmes, however, still introduce AI as an add-on productivity tool and provide few structured opportunities for pre-service teachers to analyse AI output against pedagogical and contextual demands. As a result, programmes remain insufficiently equipped to prepare beginning teachers for reflective, responsible AI integration. Without such scaffolding, AI use risks remaining instrumental – emphasising task completion over deliberation, and the development of professional judgment is left to chance.

A second, conceptual gap follows from this practical one. Much emerging research frames AI integration in terms of efficiency, idea generation, or AI literacy, and reports outcome-level improvements or perceptions of usefulness (for example, Hofmann et al., 2025; Wei et al., 2025). In comparison, sparse work has examined the cognitive and professional processes through which AI interaction becomes pedagogically meaningful. The present study addresses this gap by proposing, rather than asserting as established, that reflective processes mediate the relationship between AI-generated input and the development of learning design competency.

The objectives of the present study are therefore, threefold – first, to examine within-group change in learning design competency associated with participation in an AI-integrated reflective learning design,

second, to examine within-group change in reflective thinking, and third, to explore the reflective processes through which pre-service teachers interpret, question, modify, and reject AI-generated suggestions. These objectives are addressed by the research questions presented in the section on research questions.

LITERATURE REVIEW

This review situates the study within four bodies of scholarship that extend beyond AI-specific discourse, namely (a) established frameworks for technology integration and self-regulated learning, (b) reflective and sociocultural theories of professional learning, (c) responsible-AI literature, and (d) emerging studies of AI-supported reflection, from which the specific gap is derived.

Frameworks for Technology Integration and Self-Regulated Learning

Research on educational technology integration provides foundations that the AI-in-education literature has only begun to incorporate. The Technological Pedagogical Content Knowledge (TPACK) framework theorises teaching with technology as the interaction of technological, pedagogical, and content knowledge rather than as tool proficiency alone (Mishra & Koehler, 2006). The Substitution–Augmentation–Modification–Redefinition (SAMR) model distinguishes substitutive technology use from uses that transform learning tasks, although critics note its limited attention to context and pedagogy (Hamilton et al., 2016). Self-regulated learning theory, in turn, frames competence development as a cyclical process of forethought, performance, and self-reflection (Zimmerman, 2002). Read together, these frameworks imply that the value of generative AI for learning design depends less on the tool than on the pedagogical reasoning and self-regulatory processes surrounding its use – an implication that motivates the present study's focus on reflective mediation.

Reflective and Sociocultural Theories of Professional Learning

In teacher education scholarship, professional learning is understood as constructing, interrogating, and refining pedagogical knowledge through practice-based inquiry. Schön's (2017) account of reflection-in-action and reflection-on-action positions reflection as the mechanism through which practitioners transform experience into professional knowledge. This view is complemented by sociocultural and experiential traditions which Schön's (2017) original postulations did not address. Vygotsky's (1978) concept of mediation holds that higher-order thinking develops through culturally and materially mediated activity, suggesting that a generative-AI tool may function as a mediating artefact rather than an autonomous agent. Kolb's (1984) experiential learning cycle – concrete experience, reflective observation, abstract conceptualisation, and active experimentation, offers a complementary account of how iterative design-and-revision cycles can convert AI interaction into learning. Metacognition research further specifies the monitoring and regulation processes that such reflection entails (Flavell, 1979). Integrating these perspectives reframes reflective mediation as a theoretically grounded proposition, but still *proposes* an explanatory process, not a demonstrated mechanism.

Responsible and Ethical AI in Teacher Education

Because pre-service teachers must learn to use AI responsibly, the present review also draws on the literature on responsible AI. Policy and scholarly analyses emphasise reliability, bias, privacy, transparency, verification of outputs, and human oversight as conditions for ethical educational AI use (Holmes et al., 2022; UNESCO, 2021). Studies of pre-service teachers similarly emphasise reflective and ethical engagement with AI and a triadic view of AI literacy spanning the educator, the learner, and the tool (Heine & König, 2025; Prilop et al., 2025). For teacher education specifically, these concerns imply that AI literacy must include the disposition and capacity to verify, question, and override AI suggestions – capacities which are closely aligned with critical reflection. The present study treats responsible-AI competencies not as an external add-on, but as an outcome which reflective scaffolding is intended to cultivate.

Emerging Studies of AI-Supported Reflection

A growing body of work has examined AI-supported lesson planning and feedback (van den Berg & du Plessis, 2023; Cai et al., 2025), while some studies have reported reflective gains. For example, Wei et al. (2025) found increased post-design reflection when generative AI activities were paired with guided prompts, while Hofmann et al. (2025) reported the use of AI-generated feedback to prompt reconsideration of pedagogical assumptions. Adjacent work has framed generative AI as a collaborative, co-thinking partner (Shukla & Pandey, 2025; Woo & Cho, 2025) and connects reflective thinking with AI-supported knowledge synthesis (Ficko et al., 2025). This literature, however, has three recurring limitations that justify the present design. First, it is often descriptive rather than analytical. These studies have reported previous work to emphasise efficiency or usefulness without examining why this emphasis persisted or whether reflective dimensions were already addressed in adjacent fields, such as instructional design and self-regulated learning. Second, much of the evidence rests on short-term interventions, self-reported perceptions, or purely quantitative outcomes, which limits insight into process. Third, the literature is poorly contextualised; although AI integration is frequently treated as a universal challenge, Southeast Asian and Thai teacher-education contexts, marked by rapidly evolving digital readiness, varied institutional AI policies, and culturally specific norms of reflective practice have been sparsely examined.

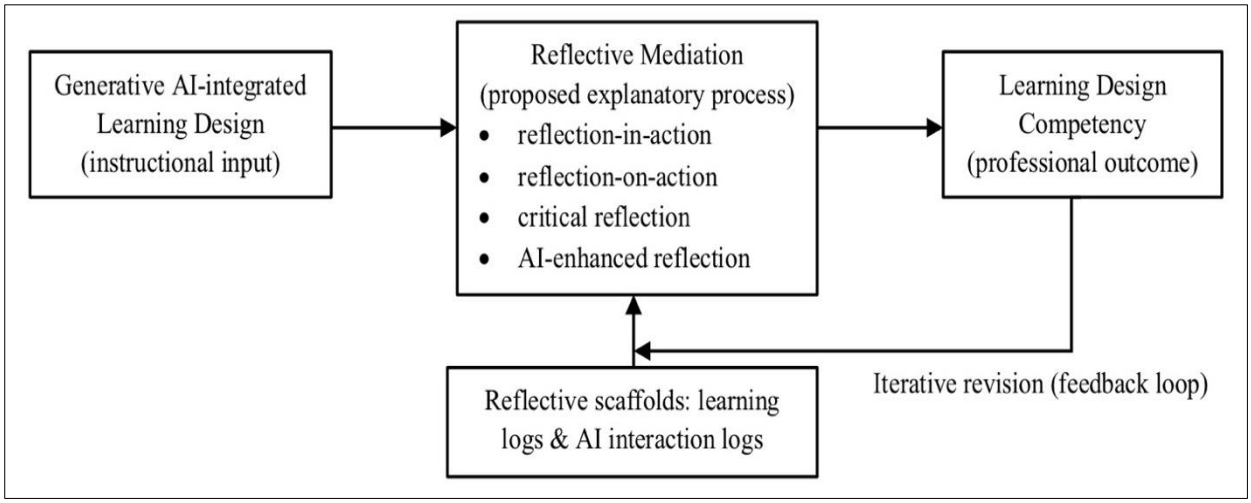
Hence, these observations sharpen the gap in the existing discussions in the field. Prior studies have already investigated AI-supported reflection, so the present contribution is not the discovery of reflection per se. Rather, what remains underexplored is a process-level account that takes into account the following: (a) integrates outcome and process evidence through a mixed-methods joint display, (b) conceptualises generative AI as a co-thinking partner whose value is contingent on reflective mediation, and (c) is grounded in established educational theory rather than just AI-specific discourse alone. The study pursues this narrower, better-evidenced gap.

CONCEPTUAL FRAMEWORK OF AI-INTEGRATED REFLECTIVE LEARNING DESIGN

This study is grounded in an integrated framework linking the following three constructs: generative AI–supported learning design, reflective thinking, and learning design competency. Rather than treating generative AI as an autonomous instructional agent, the framework positions it as a pedagogical support tool whose educational value depends on its being embedded within structured reflective processes (see Figure 1).

Figure 1

Conceptual Framework of AI-Integrated Reflective Learning Design



Generative AI as a Pedagogical Support in Learning Design

Generative AI is conceptualised as a co-thinking and feedback-support tool which assists pre-service teachers across the learning design process. Prior research indicates that it can facilitate lesson planning by generating ideas, suggesting alternatives, and providing structural feedback (van den Berg & du Plessis, 2023; Cai et al., 2025), without scaffolding, however, AI use may remain instrumental (Bae et al., 2024; Liu et al., 2025). The framework, therefore, embeds AI within structured design phases and treats AI outputs as prompts for pedagogical reasoning rather than finalised solutions.

Reflective Thinking as a Proposed Mediating Mechanism

Reflective thinking is positioned as the central *proposed* mediating mechanism. Drawing on Schön (2017), Vygotsky (1978), and Kolb (1984), reflection is understood as the process through which AI-mediated experience is examined during practice (reflection-in-action) and reconsidered afterwards (reflection-on-action) and converted into professional knowledge. However, Generative AI may introduce suggestions that do not align with learner characteristics or contextual constraints, creating a pedagogical tension that prompts reassessment, justification, and refinement. Reflective thinking in the present study encompasses reflection-in-action, reflection-on-action, critical reflection, and AI-enhanced reflection (Machost & Stains, 2023; Wei et al., 2025). Consistent with the single-group design, mediation is interpreted at a conceptual and explanatory level and is not tested statistically; it is advanced as a hypothesis for future confirmatory work rather than a demonstrated effect.

Learning Design Competency as a Professional Outcome

Learning design competency is the expected primary outcome of this study, that is, pre-service teachers' ability to design coherent instructional plans that align learner characteristics, objectives, activities, technology integration, assessment, and reflective improvement (Mishra & Koehler, 2006). This research

strategy is designed to reflect situated professional judgment rather than just technical proficiency alone (Schön, 2017), and is conceptualised as emerging through sustained reflective engagement with AI-supported design rather than through efficiency gains.

RESEARCH QUESTIONS

Consistent with the use of the non-causal, single-group design, the research questions were framed in terms of associated change and process rather than attributable effect.

RQ1. To what extent do pre-service teachers' learning design competency scores **change from pretest to posttest** following their participation in an AI-integrated reflective learning design?

RQ2. To what extent do pre-service teachers' reflective thinking scores **change from pretest to posttest**, in terms of (a) reflection-in-action, (b) reflection-on-action, (c) critical reflection, and (d) AI-enhanced reflection?

RQ3. How do pre-service teachers experience and interpret their interactions with generative AI during learning design activities within a reflective learning design?

METHOD

Research Design

This study employed a convergent mixed-methods design (Creswell & Plano Clark, 2018), in which a quantitative one-group pretest-posttest strand and a qualitative strand were collected during the same period, and merged at the interpretation stage to explain and elaborate the quantitative pattern. The design was specified as convergent with an explanatory emphasis. This means neither strand was subordinate, but the qualitative data were used to interpret how the observed quantitative change occurred. The one-group structure was shaped by authentic curricular conditions, whereby, activities were embedded within a required course in which all enrolled pre-service teachers participated as part of their regular programme, so introducing random assignment or a comparison group would have required restructuring course delivery. This design supports ecological validity but limits causal inference, and findings are reported accordingly as within-group developmental change rather than the result of intervention effects.

Participants, Selection, and Researcher-Instructor Roles

Participants were 60 pre-service teachers in their second year of an undergraduate teacher education programme at a public university in Thailand (mean age 20.8 years, SD=1.2; 24 males, 40.0%; 36 females, 60.0%). The sample was recruited through convenience sampling from a single required learning design course. All enrolled students were invited and full consent was obtained, so there was no selective enrolment. All 60 participants provided both pretest and posttest data and the full set of logs, so there was no attrition for the analysed measures. The sample comprised three subject specialisations, namely Social

Studies (n=20), Physical Education (n=20), and Dance Education (n=20). These specialisations were combined because they share the same required learning design course, the same instructional-design rubric, and the same generic (rather than subject-specific) design tasks; pooling therefore, reflects a common instructional context rather than an assumption of equivalence. Because each subgroup comprised only 20 participants, the study was not powered for subgroup comparison, and possible differences by specialisation or prior AI familiarity. These issues were treated as a limitation and an avenue for future research rather than analysed here in the present study. Familiarity with generative AI ranged from minimal to moderate, and none of the participants had received any prior formal training in an AI-integrated instructional design.

To address the potential influence of the research team on responses, the course instructor administered all coursework, whereas pretest/posttest assessments and logs were anonymised and scored independently of course grading. Participation was voluntary, written informed consent was obtained, and students were informed of their right to withdraw without academic penalty, whereby, participation, non-participation, and the content of logs had no implication on course grades. These procedures were intended to reduce demand characteristics in the self-report and reflective data.

Table 1

Demographic Characteristics of Participants (n=60)

Variable	n	%
Gender		
Male	24	40.0
Female	36	60.0
Specialisation		
Social Studies	20	33.3
Physical Education	20	33.3
Dance Education	20	33.3
Year of Study		
Second Year	60	100

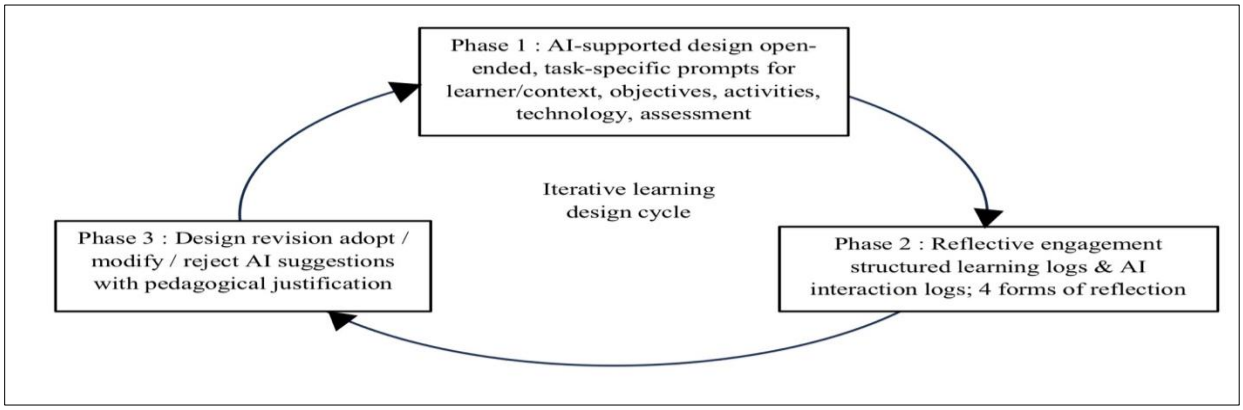
Notes. Mean age=20.8 years (SD=1.2)

Intervention: AI-Integrated Reflective Learning Design

The AI-integrated reflective learning design was embedded within the learning design course and framed generative AI as a co-thinking and feedback-support tool rather than as a content generator. To support replication, the operational structure was specified as follows. Participants engaged in iterative learning design cycles, each comprising the following three phases (see Figure 2): (1) AI-supported design, (2) reflective engagement, and (3) design revision.

Figure 2

Iterative Three-phase Structure of the AI-integrated Reflective Learning Design



In the AI-supported design phase, participants used generative AI for specific planning tasks (learner/context analysis, objectives, activities, technology integration, and assessment), using open-ended, task-specific prompts that requested explanations, alternatives, and justifications rather than complete lesson plans. In the reflective engagement phase, immediately after AI interaction, participants documented reflections in structured learning logs and AI interaction logs, guided by prompts targeting reflection-in-action, reflection-on-action, critical reflection, as well as AI-enhanced reflection. In the design revision phase, participants revised plans and explained why each AI suggestion was adopted, modified, or rejected. Operationally, the cycle was repeated across the sequence of lesson-planning tasks for the course, where each participant produced one learning log and one AI interaction log per cycle, and participants were explicitly instructed that AI outputs were not authoritative, but required verification against pedagogical principles and curricular standards.

Instruments and Validation

Learning design competency was assessed with a researcher-developed rubric covering six dimensions (learner/context analysis, learning objectives, instructional activity design, technology integration, assessment design, reflective improvement), and each was scored between 0 to 3. Content validity was established through reviews by three curriculum and instructional design specialists, and inter-rater reliability on 20% of randomly selected pretest and posttest tasks was considered strong and acceptable (ICC=.78-.85). The three experts independently reviewed every item for content relevance, clarity, and alignment with the intended construct, and items flagged as ambiguous or misaligned were revised before administering the instrument. Subsequently, only items on which the experts concurred were retained. Across both instruments, the Item-Objective Congruence (IOC) index ranged from 0.67 to 1.00, exceeding the accepted criterion of 0.50, and items scoring below this threshold were revised before administering the instrument. Reflective thinking was measured with a 12-item questionnaire (four three-item dimensions, 1-4 Likert). Internal consistency was satisfactory (overall α=.82 pretest, .86 post-test; subscales .71-.83).

The psychometric evidence has limits that the research has acknowledged transparently. Validation rested on expert review and reliability estimates, the instruments were not subjected to exploratory or confirmatory

factor analysis, and no separate large-sample pilot was conducted. This was noted as a particular constraint for the reflective thinking scale, given the theoretical multidimensionality of the reflective constructs. Accordingly, the quantitative findings are interpreted as indicative rather than definitive measurement of latent constructs (Creswell & Plano Clark, 2018), and full psychometric validation (including factor-analytic work) has been identified as a priority for future studies.

Researcher Positionality and Reflexivity

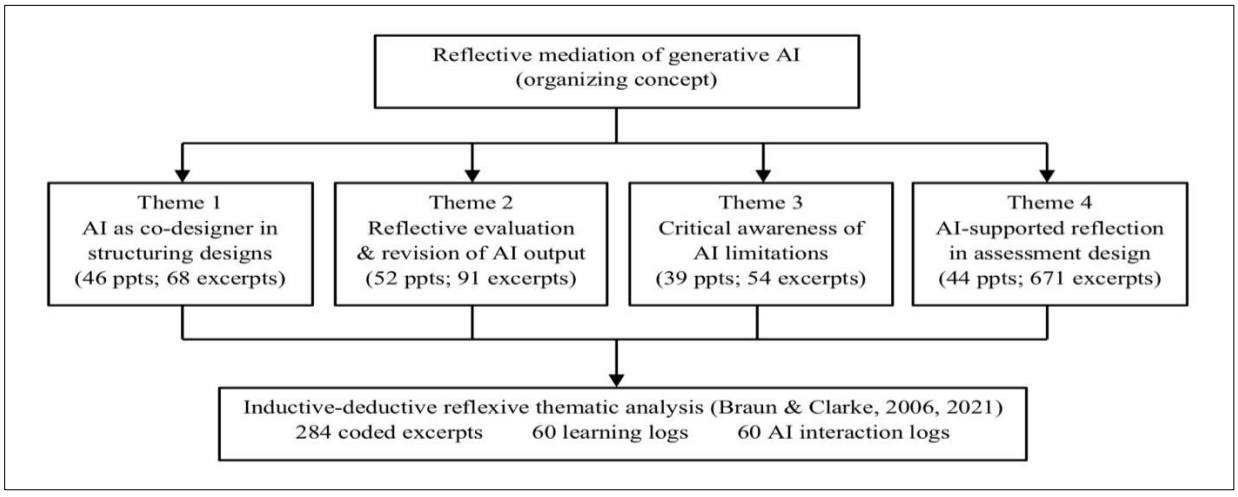
Because the study was conducted within a course, reflexivity was considered essential (Braun & Clarke, 2021). The research team comprised teacher educators with a background in instructional design and prior commitments to reflective-practice pedagogy. These researcher orientations informed both the design of the intervention and the deductive component of the coding frame and could predispose the analysis toward confirming reflective mediation. The researchers took several steps to make this stance visible and to counter it, the deductive frame was paired with inductive coding open to disconfirming evidence, instructor role was separated from independent scoring (see above), and an explicit search for negative cases (confusion, over-reliance, resistance) was built into the analysis (reported in Results). We acknowledge that our interpretive commitments shape, and are a limitation of, the qualitative claims.

Qualitative Analysis and Trustworthiness

Learning logs and AI interaction logs (284 coded excerpts across 60 of each log type) were analysed using reflexive thematic analysis following Braun and Clarke (2006, 2021). The reflexive thematic analysis proceeded through six phases, namely familiarisation, generating initial codes, constructing themes, reviewing themes, defining and naming themes, and reporting. An inductive–deductive approach allowed themes to emerge while remaining oriented to the conceptual framework (see Figure 3). All participant quotations presented in the study results were originally expressed in Thai and were translated into English by the authors.

Figure 3

Coding Framework: Four Themes Constructed Under the Organising Concept of Reflective Mediation



Notes. ppts = participants; excerpts = coded excerpts.

Trustworthiness was supported through several procedures (Lincoln & Guba, 1985). Two researchers independently coded a 20% subsample and reached substantial agreement (Cohen's κ=.81) before reconciling the full corpus. An audit trail of coding decisions and theme revisions was also maintained. Peer debriefing was conducted with a colleague not involved in the intervention; member-checking summaries of themes were shared with a subset of participants for confirmation and negative-case analysis was used to seek and report disconfirming evidence. These procedures are implemented to help strengthen credibility and dependability claims that the original manuscript left implicit.

Quantitative Analysis

Paired-samples t-tests compared pretest and posttest scores for overall and dimension-level learning design competency and reflective thinking. Effect size scores were calculated as Cohen's d_z, and 95% confidence intervals for mean differences were computed. Significance was evaluated at .05. A priori power analysis was not conducted since the sample was fixed by course enrolment. A sensitivity analysis indicates that with n=60, α=.05, and power=.80, a two-tailed paired test can detect effects of approximately $d_z = 0.37$ or larger. This is reported so readers can judge the precision of the design rather than implying post hoc power justification.

Integration of Strands

Integration occurred at interpretation through a joint display (see Table 4) that places each quantitative outcome alongside the qualitative theme that helps explain it in a merged inference. This formal procedure replaces the general narrative integration of the original manuscript and makes the mixed-methods reasoning auditable.

RESULTS

Quantitative Results

Scores increased significantly from pretest to post-test across all dimensions and overall learning design competency (see Table 2), with the largest change in technology integration and substantial change in reflective improvement.

Table 2

Pretest and Posttest Comparisons of Learning Design Competency (n=60)

Dimension	Pretest M (SD)	Posttest M (SD)	t(59)	p	d_z
Learner and context analysis	1.14 (0.26)	2.61 (0.26)	32.13	< .001	4.15
Formulation of learning objectives	2.24 (0.53)	2.60 (0.42)	4.90	< .001	0.63
Instructional activity design	2.48 (0.56)	2.66 (0.48)	3.46	.001	0.45
Technology integration	0.90 (0.39)	2.43 (0.20)	26.41	< .001	3.41
Assessment design	2.30 (0.49)	2.42 (0.44)	4.76	< .001	0.61

(continued)

Dimension	Pretest M (SD)	Posttest M (SD)	t(59)	p	d_z
Reflective improvement of learning design	1.73 (0.25)	2.47 (0.19)	18.26	< .001	2.36
Overall learning design competency	**1.80 (0.23)**	**2.53 (0.20)**	**29.04**	**< .001**	**3.75**

Notes. Effect size is Cohen's d_z. Given the bounded 0–3 rubric and low pretest scores in some dimensions, large d_z values are interpreted as substantial within-group progression rather than estimates of causal magnitude.

Reflective thinking also increased significantly overall, t(59)=9.88, p<.001, d_z=1.28 (mean increase 6.57, 95% CI [5.24, 7.90]). The largest dimension-level change was in reflection-on-action (d_z=1.51), followed by reflection-in-action (d_z=0.89), critical reflection (d_z=0.54), and AI-enhanced reflection (d_z=0.36) showed smaller change (see Table 3).

Table 3

Pretest-Post Test Comparisons of Reflective Thinking with 95% CIs (n=60)

Dimension	Pretest M (SD)	Posttest M (SD)	Mean Diff (95% CI)	t(59)	p	d_z
Reflection-in-action	2.31 (0.57)	3.11 (0.41)	0.80 [0.57, 1.03]	6.93	<.001	0.89
Reflection-on-action	2.10 (0.40)	3.08 (0.41)	0.98 [0.81, 1.15]	11.71	<.001	1.51
Critical reflection	2.32 (0.38)	2.53 (0.48)	0.21 [0.11, 0.31]	4.21	<.001	0.54
AI-enhanced reflection	2.47 (0.37)	2.66 (0.37)	0.19 [0.05, 0.33]	2.80	.007	0.36
Overall reflective thinking	27.58 (3.09)	34.15 (3.38)	6.57 [5.24, 7.90]	9.88	<.001	1.28

Notes. Mean differences are post-Pre. Effect size is Cohen's d_z. df=59 for all tests

Qualitative Results

Thematic analysis produced four themes (see Figure 3), which are presented narratively below and summarised in Table 4. Each theme is described and interpreted and subsequently illustrated with representative participant excerpts, related to Research Question 3, which probes how pre-service teachers experienced and interpreted their interactions with generative AI during the learning design activities.

Theme 1: AI as Co-Designer in Structuring Designs

The first theme, evident among 46 of the 60 participants (77%) and across 68 coded excerpts, captured the use of generative AI as a co-designer that helped participants structure their instructional plans. Participants admitted drawing on AI to generate an initial organising structure for learner and context analysis, objectives, and activities, which they later adapted to their own teaching situations. As one participant explained, *"AI gave me a starting structure, but I had to rethink the objectives to fit my students' level"* (P12). This pattern was echoed by other participants; for example, one had remarked that, *"The structure AI suggested was a good starting point, but I still had to adapt it to the context of the problem I was working*

on" (P08), while another noted, *"AI generated an initial structure for me, so I built on it and adapted it rather than copying it wholesale" (P37).* This account suggests that AI functioned as a preliminary organiser rather than a final authority, with pedagogical ownership retained by the participant. In relation to research question number 3, this theme indicates that pre-service teachers experienced AI not as a source of ready-made designs, but as a scaffold to be interrogated and localised, a disposition that parallels the strong quantitative gains in learner and context analysis, as well as technology integration.

Theme 2: Reflective Evaluation and Revision of AI Output

The second and most strongly represented theme, expressed by 52 participants (87%) across 91 excerpts, concerned the reflective evaluation and revision of AI-generated output. Participants routinely appraised AI suggestions against their lesson objectives and learners' needs, and then adopted, modified, or rejected them accordingly. One participant recounted, *"AI suggested a group discussion activity, but it did not match my lesson objective. I decided not to use it"* (P22). Similar reasoning was evident among other participants; one observed, *"Some of the AI's suggestions were usable as they were, but others had to be revised before they could be applied"* (P19), and another reflected, *"Each time I revised, I understood the problem more clearly and could improve my own work"* (P46). Such accounts demonstrate reflection-in-action and reflection-on-action in practice, as participants reasoned about the fit between AI output and pedagogical intent rather than accepting suggestions uncritically. With respect to research question number 3, the theme shows that participants experienced AI interaction as an occasion for deliberation, and it provides a qualitative counterpart to the significant quantitative gains in participants' reflection-on-action and reflection-in-action.

Theme 3: Critical Awareness of AI Limitations

The third theme, reported by 39 participants (65%) across 54 excerpts, reflected a growing critical awareness of the limitations of generative AI. Participants noted that AI output could appear plausible yet omit important curricular details, prompting some participants to regularly practise verification before use. As one participant observed, *"The explanations looked correct, but some details were missing. I decided to verify everything before using it"* (P29). Other participants expressed comparable caution; one stated, *"I did not trust the AI entirely, because some recommendations were still too general and did not match my situation"* (P16), while another concluded, *"Using AI made me realise that not every suggestion can be applied immediately without consideration"* (P32). This verification practice represents an emerging form of critical reflection and responsible-AI behaviour. In answering research question number 3, the theme reveals that participants' experience of AI included healthy scepticism, whereby its more modest representation was consistent with the comparatively smaller quantitative gains in critical and AI-enhanced reflection, suggesting that such higher-order awareness was still developing.

Theme 4: AI-Supported Reflection in Assessment Design

The fourth theme, evident among 44 participants (73%) across 71 excerpts, involved AI-supported reflection in the design of assessment. Participants reported that AI prompted them to reconsider how they measured student understanding and to weigh alternative assessment formats for feasibility and alignment. One participant reflected, *"The AI suggested alternative assessment formats, which made me rethink how*

I measure understanding" (P55). This reflective use was echoed elsewhere; one participant noted, *"AI helped me see flaws in my work that I had not noticed before"* (P28), and another explained, *"After considering the AI's suggestions, I revised several parts so that my design would be more workable in practice"* (P35). It can be seen that AI functioned as a reflective trigger that surfaced assessment decisions for deliberate evaluation rather than supplying completed instruments. With regards to research question number 3, the theme indicates that participants experienced AI as a stimulus for reflective decision-making in assessment, complementing the quantitative gains in the reflective improvement of its design.

Table 4

Summary of Qualitative Themes: Representation, Representative Excerpts, and Interpretive Meaning

Theme	Reps/excerpts	Representative excerpt	Interpretive meaning
1. AI as co-designer in structuring designs	46 (77%) / 68	"AI gave me a starting structure, but I had to rethink the objectives to fit my students' level." (P12)	AI served as a preliminary organiser adapted to context, not a final authority.
2. Reflective evaluation and revision of AI output	52 (87%) / 91	"AI suggested a group discussion activity, but it did not match my lesson objective. I decided not to use it." (P22)	Participants interrogated, modified, and rejected AI output—evidence of reflection-in/on-action.
3. Critical awareness of AI limitations	39 (65%) / 54	"The explanations looked correct, but some details were missing. I decided to verify everything before using it." (P29)	Growing critical reflection and verification disposition (responsible-AI behaviour).
4. AI-supported reflection in assessment design	44 (73%) / 71	"The AI suggested alternative assessment formats, which made me rethink how I measure understanding." (P55)	AI functioned as a reflective trigger for assessment alignment, evaluated for feasibility.

Notes. Reps=number (and %) of participants contributing to the theme. P=participant.

Disconfirming and Negative Cases

A balanced reading requires attention to participants for whom the process was difficult, and the original analysis was also under-reported. Several participants described confusion or cognitive load when AI outputs were lengthy or generic, and a small number reported a tendency towards over-reliance early in the course ("At first, I trusted the AI suggestions," P41) before make any adjustments. A few participants also expressed resistance to the reflective logging itself, finding it to be repetitive, while a few found AI suggestions to occasionally reinforce rather than challenge their initial plans. These cases indicate that reflective mediation was unbalanced across participants and that AI integration produced genuine tensions, such as over-reliance, verification burden, and reflective fatigue, which are pedagogical challenges rather than only evidence of critical reflection.

Integrated Findings

The integration of the two strands is presented narratively below and summarised in Table 5. The convergence was patterned rather than uniform. The strongest competency gains, in technology integration and learner and context analysis, aligned with Theme 1, in which participants used AI as a structural co-designer they adapted to context. The large gains from low baselines are also therefore, interpreted as structured, reflectively mediated support rather than mere efficiency. The marked gains in reflection-on-action and reflection-in-action corresponded to Theme 2, in which participants documented cycles of evaluating, modifying, and rejecting AI output. It was found that the more modest gains in critical and AI-enhanced reflection aligned with Theme 3 and with negative cases, suggesting that these higher-order capacities developed unevenly and require longer cultivation. Finally, the gains in the reflective improvement of design and in assessment design corresponded to Theme 4, in which AI acted as a reflective trigger for assessment alignment. Across the four pairings, the qualitative themes did not merely accompany the quantitative results, but helped explain how the observed changes came about, although the convergence remains associative within a single group.

Table 5

Integrated Display of Quantitative Outcomes and Qualitative Themes

Quantitative outcome	Corresponding qualitative theme	Merged inference
Large change in technology integration (d_z=3.41), learner/context analysis (d_z=4.15)	Theme 1: AI as co-designer in structuring designs	Structural AI support, evaluated and adapted, coincided with the strongest competency change, consistent with progression from low baselines.
Strong change in reflection-on-action (d_z=1.51), reflection-in-action (d_z=0.89)	Theme 2: Reflective evaluation and revision of AI output	Documented cycles of evaluating, modifying, and rejecting AI output parallel reflective-thinking gains.
Smaller change in critical reflection (d_z=0.54), AI-enhanced reflection (d_z=0.36)	Theme 3: Critical awareness of AI limitations, negative cases	Critical/AI-specific reflection developed unevenly, verification and over-reliance tensions suggest these capacities need longer cultivation.
Change in reflective improvement of design (d_z=2.36), assessment design (d_z=0.61)	Theme 4: AI-supported reflection in assessment design	AI acted as a reflective trigger for assessment alignment, evaluated for feasibility rather than adopted wholesale.

Notes. The integrated display supports interpretation, not causal mediation. Convergence is associational within a single group.

DISCUSSION

This study examined developmental change and the reflective processes associated with an AI-integrated reflective learning design. The convergence of quantitative change and documented reflective processes is consistent with, although not ascertaining the proposed view that generative AI contributes to professional learning when scaffolded through structured reflection (Schön, 2017; Vygotsky, 1978).

Interpreting the Magnitude and Pattern of Findings

The largest competency gains appeared in technology integration and learner and context analysis, the two dimensions with the lowest pretest scores. Part of this magnitude reflects a floor effect, that is, since participants began with little exposure to technology-integrated design, structured AI support had sufficient scope to increase their performance. The gain is unlikely to be a measurement artefact alone since the qualitative data show participants actively reworking AI suggestions to fit their respective learners' needs (Theme 1). This pattern echoes findings by Cai et al. (2025), who found that ChatGPT supported lesson-plan critiques improved pre-service teachers' planning skills, as well as in the study by van den Berg and du Plessis (2023), who reported that generative AI significantly assisted lesson planning. The findings depart, however, from efficiency-focused accounts: here, improvement coincided with deliberate adaptation rather than direct adoption, suggesting that the gain stemmed from pedagogical reasoning about AI output rather than from the output itself. A competing explanation warrants due acknowledgement. First, repeated exposure to iterative design tasks, rather than AI itself, may have contributed to these gains, and although participants evidently revised AI-generated suggestions, the single-group design cannot fully disentangle the contribution of repeated practice from that of AI-supported reflection.

Among the reflective dimensions, reflection-on-action was found to have the highest improvement. The most likely explanation is structural, which means that the learning and AI interaction logs required retrospective evaluation after each design cycle, directly cultivating reflection-on-action. This converges with the findings of Wei et al. (2025), who found that generative AI–assisted digital storytelling enhanced pre-service teachers' reflective thinking; and also with the results of Hofmann et al. (2025), who reported that AI-generated feedback prompted teachers to reconsider their pedagogical assumptions. Since these different designs, namely, digital storytelling, AI-generated feedback, and the present log-based reflection converge on the same result, it is postulated that reflective scaffolding, rather than generative AI itself, drives the reflective gains.

Critical reflection and AI-enhanced reflection, by contrast, was found to have the least degree of change. This is theoretically expected, since these higher-order capacities of questioning underlying assumptions and reflecting on one's own use of AI, typically take longer to develop than a single course allows (Flavell, 1979; Zimmerman, 2002). The result accords with the views of Bae et al. (2024) and Prilop et al. (2025), who reported pre-service teachers' critical and ethical engagement with AI to be uneven and emergent. The negative cases of early over-reliance and verification fatigue observed here point to a mechanism, that is, when AI output appears fluent and authoritative, the impulse to question it weakens, a tension likewise noted in studies on human–AI collaboration (Shukla & Pandey, 2025; Woo & Cho, 2025).

Situating Findings in Established Theory

The pattern is interpretable through several established frameworks that extend beyond reflective-practice theory alone. From a Vygotskian perspective (Vygotsky, 1978), generative AI operated as a mediating artefact; its value emerged not from the tool, but from how participants appropriated and regulated it, which fits the observation that the largest competency change accompanied the most structural AI support (Theme 1). Kolb's (1984) experiential cycle offers a complementary reading of the three-phase design, in which AI-supported experience, reflective observation in the logs, and active revision plausibly contributed to the

strong reflection-on-action change. The TPACK framework (Mishra & Koehler, 2006) clarifies that gains in technology integration were entangled with pedagogical and content reasoning rather than tool proficiency, while self-regulated learning theory (Zimmerman, 2002) frames the verification and override behaviours in Theme 3 as an forethought-performance-reflection regulation. Read against these theories, the present study extends, rather than merely confirms, AI-in-education work by specifying candidate mechanisms grounded outside AI-specific discourse.

Reflection emerged as the most prominent qualitative pattern, and reflection-on-action showed the largest reflective change. We are, however, deliberately cautious, as mediation was not tested statistically, and qualitative evidence alone is insufficient in establishing mediation in a causal sense. We, therefore, frame reflective mediation as a proposed explanatory account consistent with the data, instead of a demonstrated mechanism. Alternative explanations remain plausible and are not excluded by this design, including repeated lesson-planning practice, structured scaffolding independent of AI, collaborative learning, maturation, and growing familiarity with the rubric. The smaller change in critical and AI-enhanced reflection, together with negative cases, indicates that deeper critical capacities developed unevenly and may require longer engagement.

Engaging Contradictory Evidence

Rather than reframing every difficulty as critical reflection, the researchers read participants' concerns about generic outputs, missing curricular details, verification burden, and early over-reliance as genuine pedagogical and technological tensions. These tensions qualify the positive pattern, that is, AI integration imposed real costs (cognitive load, verification effort, occasional reinforcement of weak initial plans), which scaffolding mitigated, but did not remove. This more balanced interpretation tempers the novelty claim and points to design refinements, such as lighter, better-timed reflective prompts and explicit instruction in verification for future iterations.

Implications for Teacher Education

The findings carry practical implications that can be stated concretely. For curriculum design, generative AI is better embedded within reflective design cycles than introduced as a standalone tool; programmes can sequence AI-supported design, structured reflection, and justified revision across existing learning design coursework. Structured reflective scaffolds of learning logs and AI interaction logs appear central to converting AI interaction into professional learning and to developing professional judgment (Machost & Stains, 2023; Zimmerman, 2002) and should be designed to elicit verification and override decisions, not only positive appraisals. Finally, responsible-AI capacities (reliability, bias, privacy, and verification) can be cultivated through these same reflective scaffolds, rather than taught as a separate module (Holmes et al., 2022; UNESCO, 2021).

Contribution to Global Literature and Transferability

Although the study was conducted in the context of a Thai teacher education programme, it resonates with global conversations about how generative AI should be incorporated into teacher preparation. Much of the current foreign literature documents pre-service teachers' perceptions, intentions, and attitudes toward AI (Bae et al., 2024; Espino, 2025; Heine & König, 2025; Liu et al., 2025). The present findings move this

work from perception to process, showing how reflective engagement can be designed and what it yields. The study, therefore, offers a transferable design principle rather than a context-bound result; the educational value of generative AI depends less on the tool or the setting than on the reflective structures that surround its use.

This principle travels well across systems with differing resources and policies. The reflective scaffolds used here (structured learning and AI interaction logs that ask students to justify the AI suggestions they adopt, modify, or reject) are low-cost, discipline-general, and independent of advanced infrastructure, and can be built into existing coursework where dedicated AI courses or sophisticated tools are unavailable. Transfer should be theorised rather than assumed, since cultural norms of reflection, institutional AI policies, language of instruction, and learners' baseline digital readiness vary across systems and are likely to moderate both engagement and outcomes (Holmes et al., 2022; UNESCO, 2021). Comparative, multi-site research is, therefore necessary to confirm if the reflective-mediation pattern holds across educational contexts; for now, the present study is best read as a transferable hypothesis for the global agenda. Beyond initial teacher education, the same principle may also inform professional development for in-service teachers, broadening its relevance across the career continuum.

Contribution and Scope

The contribution of the study is best understood as a solid, exploratory, practice-oriented advance rather than a theoretical breakthrough. Its value lies in the following: (a) integrating outcome and process evidence through an integrated display, (b) grounding AI integration in established educational theory, and (c) offering a transferable pedagogical principle: the proposition that the educational value of generative AI may depend less on the tool than on the reflective structures surrounding its use; a proposition for which this single-group study provides preliminary rather than confirmatory support. Claims about generalisability and innovation are intentionally measured, given the single-group design and single-institution context.

LIMITATIONS AND FUTURE RESEARCH

First, it is acknowledged that the one-group pretest-posttest design constrains causal inference, maturation, testing effects, repeated practice, rubric familiarity, and concurrent instruction cannot be excluded. As such, the findings presented here are developmental and associational. Comparison conditions, such as conventional instruction, reflection without AI, and AI without reflective scaffolding are necessary to isolate the contribution of AI and of reflection. Randomised or quasi-experimental designs would allow estimation of causal magnitude while preserving ecological validity. In addition, the single-institution Thai context limits transferability, and the three pooled specialisations preclude subgroup analysis, therefore, cross-cultural, cross-institutional, and subject-specific replication is warranted. Another limitation is the researcher-developed instruments which were validated only through expert review and reliability, without factor-analytic confirmation, which future work should provide. Moreover, reflective thinking was partly self-reported and susceptible to social desirability. Therefore, the inclusion of behavioural and trace-based indicators would strengthen triangulation and help enhance the credibility and depth of findings. Finally, the absence of a delayed post-test and practicum follow-up limits conclusions about durability and transfer, which future longitudinal designs need to address.

CONCLUSIONS

This study provides preliminary, within-group evidence of participation in an AI-integrated reflective learning design which was associated with development in both learning design competency and reflective thinking, alongside a process-level account of how pre-service teachers evaluated, modified, and rejected AI-generated suggestions. The findings suggest that teacher education programmes can further transform beyond introducing generative AI as a standalone tool by embedding it within reflective cycles supported by structured logs that foster verification and professional judgment. Since the evidence obtained in this study comes from a single group in a single institution, these conclusions are offered as a basis for comparative and longitudinal confirmation rather than as definitive claims of effectiveness.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

ACKNOWLEDGMENT

This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

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

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328-352. https://doi.org/10.1080/14780887.2020.1769238

Cai, H., Han, B., Sun, J., Li, X., & Wong, L. H. (2025). Harnessing AI for teacher education to promote inclusive education: Investigating the effects of ChatGPT-supported lesson plan critiques on the development of pre-service teachers' lesson planning skills. *The Internet and Higher Education, 67*, 1-15. https://doi.org/10.1016/j.iheduc.2025.101022

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Espino, C. (2025). Pre-service teachers' perspectives on generative AI: Insights, applications, and challenges in the classroom. *International Journal on Studies in Education, 7*(3), 732-747. https://doi.org/10.46328/ijonse.1930

Ficko, A., Sarkki, S., Gultekin, Y. S., Egli, A., & Hiedanpää, J. (2025). Reflective thinking meets artificial intelligence: Synthesizing sustainability transition knowledge in left-behind mountain regions. *Geography and Sustainability, 6*(1), 1-11. https://doi.org/10.1016/j.geosus.2024.100257

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends, 60*(5), 433-441. https://doi.org/10.1007/s11528-016-0091-y

Heine, S., & König, J. (2025). Applying artificial intelligence in teacher education: Preservice teachers' attitudes and reflections in using ChatGPT for teaching and learning. *European Journal of Teacher Education, 48*(5), 934-9363. https://doi.org/10.1080/02619768.2025.2540791

Hofmann, F., Daunicht, T. M., Plößl, L., & Gläser-Zikuda, M. (2025). Promoting reflection skills of pre-service teachers – The power of AI-generated feedback. *Education Sciences, 15*(10), 1-17. https://doi.org/10.3390/educsci15101315

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy, and the rule of law*. Council of Europe.

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

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Liu, Y., Wang, Q., & Lei, J. (2025). Adopting generative AI in future classrooms: A study of preservice teachers' intentions and influencing factors. *Behavioral Sciences, 15*(8), 1-23. https://doi.org/10.3390/bs15081040

Machost, H., & Stains, M. (2023). Reflective practices in education: A primer for practitioners. *CBE—Life Sciences Education, 22*(2), 1-11. https://doi.org/10.1187/cbe.22-07-0148

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

Prilop, C. N., Mah, D.-K., Jacobsen, L. J., Hansen, R. R., Weber, K. E., & Hoya, F. (2025). Generative AI in teacher education: Educators' perceptions of transformative potentials and the triadic nature of AI literacy explored through AI-enhanced methods. *Computers and Education: Artificial Intelligence, 9*, 1-15. https://doi.org/10.1016/j.caeai.2025.100471

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge. https://doi.org/10.4324/9781315237473

Shukla, H., & Pandey, K. (2025). Human–AI collaboration in teaching and learning. *International Journal of Science and Social Science Research, 2*(4), 367-375. http://dx.doi.org/10.2139/ssrn.5198554

UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO. https://doi.org/10.54675/PCSP7350

van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences, 13*(10), 1-12. https://doi.org/10.3390/educsci13100998

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

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, 203(105356),* 1-21. https://doi.org/10.1016/j.compedu.2025.105356

Woo, H., & Cho, Y. Y. (2025). Human–AI collaboration: Students' changing perceptions of generative artificial intelligence and active learning strategies. *Sustainability, 17*(18), 1-22. https://doi.org/10.3390/su17188387

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2