education sciences

MDPI

*Article*

# Exploring Elementary Pre-Service Teachers' AI Literacy in ChatGPT-Assisted Science Lesson Design

**Yoon-Sung Choi [1] and Min-Cheol Kim [2,*]**

[1] Department of Science Education, Cheongju National University of Education, Cheongju 28690, Republic of Korea
[2] Department of Earth Science Education (Geology Division), College of Education, Pusan National University, Busan 46241, Republic of Korea
* Correspondence: igumany@pusan.ac.kr

**Abstract**

This study examined elementary pre-service teachers' artificial intelligence (AI) literacy in ChatGPT-4o-assisted science lesson design. Eight third-year pre-service teachers at a national university of education in Korea participated in a voluntary, non-credit, ten-week program focused on elementary Earth science lessons and used GPT-4o through the web-based ChatGPT interface. Data included pre-task written reflections, ChatGPT interaction logs, lesson plans and instructional artifacts, reflective journals, observation notes, and final reflective reports. A framework-guided thematic analysis examined AI knowledge, AI skills, AI ethics, and AI attitudes, while participant-level trajectories and lesson-design artifacts were analyzed across data sources. Initial AI literacy was heterogeneous rather than uniformly limited. Five of the eight participants articulated some pattern-based, probabilistic, or context-sensitive understanding of ChatGPT, seven of the eight anticipated using prompts that specified instructional conditions, all eight recognized a need to verify AI-generated content, and initial attitudes ranged from affirmative willingness to predominant skepticism. During the program, six of the eight participants showed increasingly contextualized or iterative prompting, and seven of the eight demonstrated artifact-linked verification, revision, supplementation, or alternative tool selection. Attitudinal trajectories were characterized by conditional acceptance rather than a uniform increase in trust. Across the eight participant-level trajectories, the analysis identified refinement, continuity, limited change, and one episode-specific awareness–practice inconsistency. These findings indicate that ChatGPT-assisted science lesson design can provide a practice-based context for examining AI literacy as situated professional judgment involving scientific, curricular, pedagogical, and ethical decision-making, without implying a common developmental progression.

**Keywords:** AI literacy; ChatGPT; generative artificial intelligence; pre-service teachers; science lesson design; teacher education

Check for updates

Academic Editor: Michael DeBraga

Received: 28 May 2026
Revised: 4 August 2026
Accepted: 6 August 2026
Published: 10 August 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

The emergence of generative artificial intelligence (GenAI) technologies is reshaping how instructional design is conceptualized and practiced in education. Among these tools, ChatGPT, a large language model that generates context-sensitive responses, has gained attention as a resource for producing explanations, suggesting instructional activities, refining lesson content, and preparing assessment tasks (Cai et al., 2025; Holmes & Tuomi,

2022; Joksimovic et al., 2023). In science education, where conceptual accuracy, inquiry-oriented pedagogy, and student-accessible representations are essential, GenAI may offer useful support for pre-service teachers as they prepare and revise instructional materials.

Meaningful use of GenAI in science teaching, however, requires more than access to a tool or the ability to write prompts. Science lesson design involves transforming disciplinary ideas into explanations, inquiry activities, models, representations, and assessments that are scientifically accurate, curriculum-aligned, developmentally appropriate, and feasible in classrooms. These demands are especially visible in elementary Earth science, where topics such as weather, seasonal change, planetary characteristics, constellations, and changes in the Moon's appearance require teachers to interpret abstract or spatial phenomena for young learners. Evaluating AI-generated content therefore draws on pedagogical and content-related judgment as well as technical skill. It also raises ethical questions concerning reliability, privacy, bias, transparency, attribution, and teacher responsibility (Cooper, 2023; Zhai et al., 2024).

Recent studies have examined how ChatGPT and other GenAI tools support science lesson planning, 5E-based instructional design, simulations, and instructional material generation (Choi, 2025b; Lee & Zhai, 2024; Okulu & Muslu, 2024; Peikos & Stavrou, 2025). These studies indicate that GenAI can provide ideas and scaffolds while also requiring contextual adaptation, verification, and professional judgment. Much of the existing research, however, has focused on what ChatGPT can generate, how users perceive its usefulness, or how it functions as an instructional design aid. Less attention has been paid to how pre-service teachers' artificial intelligence (AI) literacy becomes visible through repeated decisions to accept, revise, reject, or supplement AI-generated content during subject-specific lesson design.

This study addresses that gap by exploring elementary pre-service teachers' AI literacy in ChatGPT-assisted science lesson design. AI literacy is understood as a multidimensional capacity involving AI knowledge, AI skills, AI ethics, and AI attitudes (Ng et al., 2021; Wang et al., 2023; Yue et al., 2026). Rather than assuming a uniform developmental progression, the study examines participants' initial manifestations and potentially uneven changes during a ten-week instructional design program. Drawing on pre-task reflections, ChatGPT interaction logs, lesson plans and instructional artifacts, reflective journals, observation notes, and final reflective reports, the analysis traces how participants formulated prompts, evaluated outputs, revised materials, and positioned themselves toward GenAI. The following research questions guide the inquiry:

1. What initial manifestations of AI literacy do elementary pre-service teachers show before engaging in ChatGPT-assisted science lesson design?
2. What observable and context-specific changes in AI literacy are identified as elementary pre-service teachers engage in ChatGPT-assisted science lesson design?

## 2. Literature Review

This section reviews two bodies of literature that frame the present study. First, it examines how GenAI has been discussed in science education, science lesson design, and pre-service teacher education. Second, it reviews the concept of AI literacy and explains why practice-based teacher education can provide a meaningful context for examining pre-service teachers' AI literacy.

### *2.1. GenAI in Science Lesson Design and Teacher Education*

GenAI has become an increasingly important topic in educational research because of its capacity to generate explanations, summarize information, provide feedback, and support the development of instructional materials. In science education and teacher

education, these affordances are particularly relevant because teachers are required to transform disciplinary knowledge into student-accessible explanations, inquiry activities, assessment tasks, and instructional representations.

#### 2.1.1. GenAI in Science Education and Lesson Design

Recent advances in GenAI have opened new possibilities for supporting science education across grade levels. Tools such as ChatGPT have shown potential to assist learners and teachers in tasks such as conceptual clarification, scientific explanation, scientific writing, inquiry-based reasoning, and problem-solving (Anik et al., 2025; Biswas, 2023; Chang et al., 2023; Cooper, 2023; Joksimovic et al., 2023). These tools can provide immediate responses and diverse representations of scientific ideas, which may help learners engage with complex science content and support teachers in preparing instructional materials.

The potential of GenAI is closely related to science lesson design. Science lesson design requires teachers to organize learning objectives, select appropriate scientific content, construct inquiry activities, design explanations, and consider how students may understand scientific phenomena. In this process, GenAI tools may support teachers by generating lesson ideas, suggesting classroom activities, proposing assessment items, and offering alternative explanations for scientific concepts (Clark et al., 2024; Okulu & Muslu, 2024; Peikos & Stavrou, 2025). Such support may be particularly useful in elementary Earth science education, where teachers often need to represent abstract, spatial, and systems-based phenomena, including seasonal changes, weather systems, planetary characteristics, constellations, and changes in the Moon's appearance. These topics require instructional materials that are scientifically accurate, age-appropriate, curriculum-aligned, and connected to students' everyday experiences.

At the same time, the use of GenAI in science education raises important concerns. AI-generated responses may include scientific inaccuracies, oversimplified explanations, inappropriate analogies, or content that does not align with curriculum goals. Prior studies have emphasized that scientific content generated by GenAI should be carefully examined in terms of accuracy, reliability, and educational appropriateness (Cooper, 2023; Zhai et al., 2024). Studies in Korean science education contexts have also examined the scientific validity and instructional appropriateness of ChatGPT-generated responses in elementary science and Earth science education, suggesting that AI-generated scientific content should not be used directly without critical review and pedagogical judgment (Choi, 2024; Choi & Kim, 2025). These findings indicate that GenAI can be useful for science lesson design, but its educational value depends on teachers' ability to evaluate, revise, and contextualize AI-generated outputs.

#### 2.1.2. GenAI in Pre-Service Teacher Education

Research on GenAI in teacher education has also begun to expand. Recent studies suggest that pre-service teachers may use GenAI tools for academic tasks, study support, writing assistance, or general information seeking, but their use of GenAI for lesson planning and instructional design remains limited or uneven (Ayanwale et al., 2025; Brandão et al., 2024; Hsu et al., 2024; Okulu & Muslu, 2024; Sanusi et al., 2024). This suggests a gap between general familiarity with AI tools and the capacity to use them meaningfully for pedagogical purposes.

Several studies have emphasized that effective GenAI use in teacher education requires more than access to AI tools (Gatlin, 2023; Kotsis, 2024). Pre-service teachers need structured opportunities to evaluate, revise, and reflect on AI-generated content in realistic instructional contexts (Li et al., 2024; Purnama et al., 2023). Without such opportunities, GenAI may be used mainly for surface-level assistance, such as generating ideas or

completing academic tasks, rather than for deeper instructional reasoning. This issue is particularly important in science education because pre-service teachers must evaluate the scientific accuracy, conceptual clarity, curriculum alignment, and learner appropriateness of AI-generated content (Choi & Chung, 2025).

Lesson planning is a core component of teacher preparation because it requires pre-service teachers to synthesize content knowledge, align instructional goals with student needs, and make pedagogical decisions (Okulu & Muslu, 2024). Integrating GenAI into this process may enhance efficiency, but it also introduces new challenges and responsibilities. When pre-service teachers use ChatGPT during lesson design, they must decide whether AI-generated explanations are accurate, whether proposed activities are pedagogically meaningful, and whether the materials are appropriate for students' developmental levels. These evaluation and decision-making processes reflect deeper aspects of teacher cognition, ethics, and professional judgment (Yue et al., 2022; Zhai, 2021).

However, much of the existing research has tended to approach ChatGPT mainly as an instructional design tool or as a source of scientific responses. Previous studies have often focused on what ChatGPT can generate, how it can support lesson planning, or what limitations appear in AI-generated content. Although these approaches are important, they do not fully explain how pre-service teachers' AI literacy is manifested as they repeatedly interact with GenAI during lesson design. From a teacher education perspective, lesson design is not merely a technical task but a professional learning context in which pre-service teachers' understanding, judgment, responsibility, and attitudes toward AI may become visible.

### *2.2. AI Literacy in Practice-Based Teacher Education*

As GenAI becomes increasingly integrated into education, AI literacy has emerged as a critical focus in both AI education and teacher preparation. For pre-service teachers, AI literacy is especially important because they are not only future users of AI technologies but also future educators who must make pedagogical, ethical, and professional decisions about the use of AI in classrooms.

#### 2.2.1. Conceptualizing AI Literacy

AI literacy generally refers to the capacity to understand AI technologies, use AI tools effectively, evaluate AI-generated outputs, and make informed decisions about the ethical and social implications of AI (Ng et al., 2021). It is not limited to technical knowledge or operational skills. Rather, AI literacy involves the ability to critically engage with AI systems, recognize their limitations, and use them responsibly in social and educational contexts (Biagini, 2025; Hur, 2025).

Recent studies have conceptualized AI literacy as a multidimensional construct. AI literacy has been described as involving the ability to know and understand AI, apply AI, evaluate AI applications, and consider AI ethics (Ng et al., 2021). It has also been measured as users' competence in understanding, using, and evaluating AI (Wang et al., 2023). In addition, AI literacy has been discussed as a broad construct that includes multiple components, learning methods, and effects (Pinski & Benlian, 2024). These studies suggest that AI literacy cannot be reduced to knowing how to operate AI tools; rather, it includes cognitive, practical, ethical, and affective dimensions.

In this study, AI literacy is conceptualized through four dimensions: AI knowledge, AI skills, AI ethics, and AI attitudes. This framework is consistent with recent multidimensional approaches to AI literacy. A four-dimensional framework of AI literacy has been proposed to include AI knowledge, AI skills, AI ethics, and AI attitudes in AI education (Yue et al., 2026). This perspective emphasizes that learners should understand basic AI concepts, apply AI

tools to solve problems, consider ethical and social issues, and develop informed attitudes toward AI. Similarly, AI literacy has been shown to involve not only knowledge acquisition but also affective and cognitive engagement, including motivation, self-efficacy, confidence, ethical awareness, and higher-order thinking (Ng et al., 2024). These perspectives indicate that AI literacy includes what learners know about AI, how they use it, how they evaluate it, and how they emotionally and ethically position themselves toward it.

For pre-service teachers, AI literacy has particular significance because they need to decide when, why, and how AI should be used in classrooms (Ayanwale et al., 2024). Teacher education should therefore support AI literacy not simply as a set of technical skills but as a form of professional judgment. Pre-service teachers need to evaluate whether AI-generated explanations are scientifically accurate, whether suggested activities align with curriculum goals, whether materials are appropriate for students' developmental levels, and whether AI use raises ethical concerns related to reliability, bias, transparency, or teacher responsibility. These decisions require reflective and responsible pedagogical reasoning.

#### 2.2.2. Practice-Based Contexts for Examining AI Literacy in Pre-Service Teacher Education

Practice-based learning environments are important for examining AI literacy because they allow learners to interact with AI tools, solve problems, evaluate outcomes, and reflect on ethical and social implications. Previous AI literacy research has explored curriculum design, educational games, tool-based learning, project-based learning, and design-based learning as ways to support learners' understanding and use of AI (Ng et al., 2024; Zhang & Zhang, 2024). These approaches suggest that AI literacy can become visible through structured, interactive, and reflective experiences rather than through conceptual instruction alone.

Design-based and tool-based approaches are particularly relevant to AI literacy in teacher education. Design-based learning can support AI knowledge, AI skills, AI ethics, and AI attitudes by engaging learners in the design of AI-supported solutions for real-world problems (Yue et al., 2026). Structured and interactive learning environments can also enhance learners' motivation, self-efficacy, confidence, AI knowledge, and ethical awareness (Ng et al., 2024). Although many of these studies have focused on K–12 students, they provide useful implications for teacher education because they show that AI literacy is closely related to learners' active engagement with AI-related tasks.

In the context of pre-service teacher education, instructional design can serve as a practice-based context in which AI literacy is demonstrated and examined. When pre-service teachers design science lessons with ChatGPT, they do not simply receive AI-generated outputs. They need to formulate prompts, interpret responses, evaluate scientific accuracy, revise materials, and decide whether the outputs are appropriate for elementary students (Choi, 2025a). Through these activities, aspects of AI knowledge, AI skills, AI ethics, and AI attitudes may become visible. AI knowledge may be reflected in how pre-service teachers understand ChatGPT's generative responses and limitations. AI skills may be reflected in how they refine prompts and adapt AI-generated content to lesson goals. AI ethics may be reflected in how they consider accuracy, reliability, appropriateness, and teacher responsibility. AI attitudes may be reflected in how they move between uncertainty, anxiety, curiosity, cautious confidence, or critical engagement with AI.

Therefore, ChatGPT-assisted science lesson design can be understood as a meaningful context for examining the initial manifestations and observable changes of AI literacy in pre-service teacher education. This study builds on previous research on GenAI in science education and AI literacy education but shifts the focus from ChatGPT as a tool to the ways in which pre-service teachers engage with, evaluate, and adapt AI-generated content during science lesson design. By examining AI knowledge, AI skills, AI ethics, and AI

attitudes together, this study aims to provide an integrated understanding of how AI literacy can be observed in practice-based teacher education contexts.

## 3. Materials and Methods

### 3.1. Research Design

This study adopted a qualitative research design to examine initial manifestations and observable, context-specific changes in elementary pre-service teachers' AI literacy during ChatGPT-assisted science lesson design. A qualitative approach was appropriate because the study sought to interpret participants' instructional reasoning and interactions with GenAI, rather than quantify outcomes or test causal effects (Creswell & Poth, 2016). The inquiry was situated in a voluntary ten-week program conducted from June to August 2025, in which participants used GPT-4o through the web-based ChatGPT interface to design Earth science lessons.

The study was guided by four dimensions of AI literacy: AI knowledge, AI skills, AI ethics, and AI attitudes. These dimensions informed the research instruments, program activities, and analytical categories. AI literacy was not treated as a fixed outcome or uniform developmental sequence. Instead, the study examined how its dimensions became visible, remained stable, or changed unevenly across pre-task reflections, ChatGPT interaction logs, lesson plans and artifacts, reflective journals, observation notes, and final reflective reports. This design positioned AI literacy as a situated, practice-based phenomenon involving engagement with content-related, pedagogical, and ethical judgment. Table 1 summarizes the research design.

**Table 1.** Overview of the study design.

| Component | Description |
| --- | --- |
| Research focus | Initial manifestations and observable, context-specific changes in elementary pre-service teachers' AI literacy |
| Participants | Eight third-year elementary pre-service teachers at a national university of education in Korea |
| Program context | Voluntary ten-week ChatGPT-assisted science lesson-design program conducted from June to August 2025 |
| AI system | GPT-4o accessed through the web-based ChatGPT interface |
| Lesson topics | Seasonal Changes; Weather and Our Life; and Solar System and Stars, including planetary characteristics, constellations, and changes in the Moon's appearance |
| Data sources | Pre-task written reflections, ChatGPT interaction logs, lesson plans and instructional artifacts, reflective journals, observation notes, and final reflective reports |
| Analytic framework | AI knowledge, AI skills, AI ethics, and AI attitudes, supplemented by a criterion-guided review of lesson-design artifacts |

### 3.2. Research Procedures and Participants

Eight third-year elementary pre-service teachers at a national university of education in Korea voluntarily participated in the study. The group comprised five women and three men. All participants had completed introductory coursework in elementary science content and elementary science teaching methods and had previously designed at least one elementary science lesson. Before the program, all eight participants reported that they had no prior experience using GenAI for educational purposes, including elementary science lesson planning, instructional material development, or classroom application. Their general familiarity with GenAI varied, but none had previously applied it in an educational context. This preparation provided a foundational basis for considering scientific accuracy, curriculum expectations, learner characteristics, and instructional feasibility when

evaluating ChatGPT outputs. However, pedagogical content knowledge was not independently assessed or treated as an analytical construct; accordingly, the study does not make claims about changes in participants' PCK. To protect confidentiality, participants were assigned anonymized identifiers. During the comprehensive participant-level reanalysis, the identifiers were reassigned and cross-checked across the integrated dataset to ensure that each participant's pre-task reflection, ChatGPT interaction logs, reflective journals, lesson plans and instructional artifacts, and final reflective report were consistently linked. Accordingly, the T1–T8 identifiers used throughout this manuscript represent the final anonymized coding scheme established during the reanalysis. This reassignment concerned only the identifiers used for reporting and did not alter the underlying participant statements or data.

The first author facilitated a voluntary, non-credit, ten-week research-based program conducted outside participants' regular coursework. Participation in or withdrawal from the study had no effect on academic credit, grades, or formal evaluation. Within this program, participants used their own paid personal ChatGPT accounts, which they individually owned and controlled. No accounts or institutional licenses were provided by the researchers or the participating institutions, and the individual participants bore their own subscription costs without reimbursement from the research project. All participants had access to GPT-4o and were instructed to use this model consistently throughout the ten-week program. Chat history was retained so that the complete interaction records remained available, while the setting allowing conversations to be used for model improvement was disabled. The memory function was not used as part of the research activities. Participants were explicitly instructed not to enter personal, confidential, or identifying information into ChatGPT, and neither participant-identifiable information nor elementary pupil data were entered into the system. ChatGPT was selected because it was a widely used and readily accessible GenAI platform with a familiar interface, which helped reduce barriers to participation. Within ChatGPT, GPT-4o was specified to minimize variation arising from frequent model updates during the study period. Participants used only the standard GPT-4o chat interface and entered text-based prompts. Web browsing, uploaded reference materials, custom GPTs, external knowledge bases, retrieval-augmented generation (RAG), and other RAG-enabled functions were not used. Curricular materials and external sources were consulted separately for verification purposes but were not uploaded to or retrieved through ChatGPT. Neither author nor either affiliated institution had any financial, contractual, research, licensing, advisory, or other relationship with OpenAI in connection with this study beyond ordinary use of the publicly available ChatGPT platform.

The program provided repeated opportunities to formulate and revise prompts, examine GPT-4o responses, compare outputs with curricular or external sources, adapt lesson ideas and materials, discuss decisions with peers, conduct microteaching, and complete structured reflections. The complete set of lesson-design tasks represented in the analyzed dataset addressed Seasonal Changes, Weather and Our Life, and Solar System and Stars, including planetary characteristics, constellations, and changes in the Moon's appearance. Across the program, participants moved from initial reflection and lesson ideation to iterative AI interaction, evaluation, instructional revision, microteaching, and final reflection. These activities were intended to make participants' existing and emerging AI literacy practices observable within subject-specific instructional decision-making rather than to deliver a standardized AI-skills intervention. Appendix A provides a detailed program sequence.

### 3.3. Data Sources and Collection

Multiple qualitative data sources were collected across the ten-week program to capture participants' initial understandings, in-process instructional decisions, observable uses

of ChatGPT, and final reflections. The dataset comprised eight pre-task written reflections, sixteen reflective journals, eight final reflective reports, eight sets of ChatGPT interaction logs, eight participant-level sets of lesson plans and instructional artifacts, and researcher observation notes. The eight interaction-log sets contained 150 documented prompt–response episodes (T1–T8: 23, 16, 15, 17, 24, 16, 18, and 21, respectively). Participant-level counts reported in the Results indicate whether a practice was documented at least once for a participant, rather than how frequently or consistently it occurred across all episodes. Each participant produced and submitted an individual dataset. Before analysis, names and identifying details were removed, and participants were represented by the codes T1–T8. Because the program involved simulated lesson design and peer microteaching rather than instruction with elementary pupils, no elementary student data were collected or entered into ChatGPT.

Pre-task reflections documented initial AI knowledge, anticipated prompting strategies, ethical concerns, and attitudes toward GenAI. Interaction logs preserved participants' prompts, GPT-4o responses, follow-up questions, and written explanations of why outputs were accepted, revised, supplemented, or rejected. Lesson plans, presentation slides, worksheets, and related instructional materials provided evidence of how these decisions were incorporated into final lesson designs. Two reflective journals captured ongoing evaluations of scientific accuracy, curriculum alignment, developmental appropriateness, usefulness, and difficulties encountered during the design process. Observation notes recorded peer discussions, uncertainties, feedback, and decision-making episodes. Final reflective reports elicited participants' retrospective accounts across AI knowledge, skills, ethics, and attitudes. For the first research question, pre-task reflections served as the primary source for constructing participants' initial AI literacy profiles, with early interaction records used as supplementary evidence when clarification was needed. For the second research question, pre-task, process, artifact, and post-program data were connected at the participant level to trace continuity, refinement, observable change, inconsistency, and limited change over time. Appendix B presents the data sources, representative prompts, and analytical purposes.

### 3.4. Data Analysis

The analysis combined a framework-guided thematic analysis of AI literacy with a criterion-guided qualitative review of lesson-design artifacts. The analytical process was organized around the two research questions and connected participants' written accounts with their observable interactions, revision decisions, and final instructional materials. Rather than treating the four AI literacy dimensions as isolated categories, the analysis examined how knowledge, skills, ethics, and attitudes were combined within participants' initial profiles and subsequent lesson-design trajectories. The alignment among the research questions, analytical phases, evidence sources, analytical comparisons, and outputs is summarized in Table 2.

**Table 2.** Alignment of research questions, analytical phases, and evidence sources.

| Research Question and Analytical Phase | Primary Evidence | Analytical Comparison | Analytical Output |
| --- | --- | --- | --- |
| RQ1: Initial-profile analysis | Pre-task reflections; early interaction records when clarification was needed | Within-participant integration of four AI literacy dimensions, followed by cross-case comparison | Shared and contrasting initial AI literacy profiles |

**Table 2.** *Cont.*

| Research Question and Analytical Phase | Primary Evidence | Analytical Comparison | Analytical Output |
| --- | --- | --- | --- |
| RQ2: Within-case trajectory analysis | Pre-task reflections, interaction logs, journals, observation notes, artifacts, and final reflections | Temporal and cross-source comparison within each participant | Change, refinement, continuity, inconsistency, and limited-change trajectories |
| RQ2: Cross-case pattern analysis | Eight participant trajectories and participant-by-theme matrix | Comparison across participants and data sources | Integrated patterns and discrepant cases |
| RQ2: Artifact-based review | Interaction logs and final instructional artifacts | Prompt–response–evaluation–revision–artifact sequence | Observable instructional revisions and qualifying evidence |

#### 3.4.1. Initial-Profile and Trajectory Analysis of AI Literacy

All textual and observational data were analyzed through familiarization, initial coding, theme development, theme review, theme definition, and interpretation (Braun & Clarke, 2006). The first author repeatedly read the data and prepared analytic memos concerning participants' understandings of AI, prompting practices, evaluation and verification decisions, ethical concerns, and attitudes toward GenAI. Units of analysis ranged from individual statements to connected sequences involving a prompt, GPT-4o response, participant evaluation, follow-up request, revision decision, and related instructional artifact.

Initial coding was deductively organized around four dimensions of AI literacy. AI knowledge concerned participants' explanations of how ChatGPT generated responses and their recognition of its capabilities and limitations. AI skills concerned prompt formulation, follow-up questioning, output adaptation, and application to instructional purposes. AI ethics concerned reliability, scientific and representational accuracy, developmental appropriateness, privacy, attribution, bias, access, proprietary-platform dependence, overreliance, and teacher responsibility when these issues appeared in the data. AI attitudes concerned trust, anxiety, hesitation, curiosity, confidence, skepticism, conditional acceptance, and intended future use. Inductive subthemes were subsequently developed within and across these dimensions. Ethical issues not substantively evidenced in the dataset were not inferred as participant findings.

The analysis then proceeded in two connected phases corresponding to the research questions. In the first phase, participants' pre-task reflections were coded across the four dimensions to construct an initial AI literacy profile for each participant. For the AI-knowledge component, an explanation was classified as pattern-based, probabilistic, or context-sensitive when it explicitly referred to generation from learned patterns, prediction of likely outputs, or adjustment of responses according to the user's prompt or conversational context. Mixed accounts were included when at least one of these criteria was met, while inaccurate or incomplete assumptions were retained as qualifying evidence. These profiles represented combinations of understandings, intended prompting practices, ethical concerns, and attitudes rather than standardized levels or scores. Participant-level profiles were compared to identify shared and contrasting initial manifestations across the eight cases. Early interaction records were consulted only when they clarified an initially stated understanding or intended practice.

In the second phase, a chronological within-case trajectory was constructed for each participant by linking pre-task statements with subsequent prompts, GPT-4o responses, participant evaluations, follow-up requests, revision decisions, reflective journals, instructional artifacts, and final reflections. Claims of observable change required temporal or

cross-source evidence and were not based solely on a single retrospective statement. The analysis distinguished among newly observable practices, refinement of existing practices, continuity, awareness–practice inconsistency, limited change, and cases in which comparatively explicit initial understandings left less scope for visible change.

The eight participant trajectories were subsequently compared through a participant-by-theme matrix. The matrix identified which participants and data sources supported each cross-case pattern, where evidence converged, and where self-reports differed from interaction records or final artifacts. Participant counts were used only to describe the distribution of patterns within the eight-participant dataset and were not treated as estimates of prevalence or as grounds for statistical generalization. For example, statements referring to the prediction of likely outputs were initially coded as predictive generation, categorized within AI knowledge, and incorporated into participant-level interpretations of probabilistic understanding.

#### 3.4.2. Lesson-Design Artifact Analysis

Lesson plans, presentation slides, worksheets, and related materials were examined as artifact-based evidence within participants' trajectories rather than as a separate dimension of AI literacy. Each artifact set was reviewed together with the corresponding interaction logs to trace the sequence from an initial prompt and GPT-4o response to the participant's evaluation, follow-up request, revision or supplementation, and final instructional decision.

The review considered six criteria: scientific accuracy, curriculum alignment, developmental appropriateness, inquiry orientation, classroom feasibility, and pedagogical coherence among objectives, activities, representations, and assessment. Analysis focused on observable revisions such as correcting or verifying scientific content, narrowing curricular scope, modifying explanations or representations, restructuring activities, selecting an alternative digital tool, and adjusting materials for classroom implementation.

Artifact evidence was used to corroborate, qualify, or challenge patterns identified in participants' reflections and interaction records. For example, a stated concern about verification was examined against whether the participant actually checked, revised, supplemented, or retained an AI-generated response. The review did not generate standardized lesson-quality scores, rank participants, or evaluate classroom effectiveness or student learning. Trustworthiness procedures for the initial-profile, trajectory, cross-case, and artifact analyses are described in Section 3.5.

### *3.5. Researcher Positionality, Reflexivity, and Trustworthiness*

The first author designed and facilitated the ten-week program, collected the data, and conducted the initial coding and participant-level analyses. This dual role provided contextual familiarity but also created a potential power asymmetry and risk of socially desirable responses. Participants were informed that participation was voluntary, withdrawal involved no penalty, and either decision would not affect academic credit, grades, or formal evaluation. All materials were de-identified before analysis.

Reflexivity and trustworthiness were supported throughout the initial-profile, within-case trajectory, cross-case pattern, and artifact analyses. The first author maintained analytic memos documenting interpretive decisions, changes in code boundaries, competing explanations, and cases that did not fit the dominant patterns. The participant-by-theme matrix enabled systematic comparison of participants and data sources and helped identify continuity, limited change, and inconsistencies between reported awareness and observable practice.

All participant data were originally produced in Korean. Quotations selected for presentation in the manuscript were translated into English by the first author and cross-

checked by the corresponding author against the original Korean data. The authors discussed any differences in wording and revised the translations by consensus, prioritizing the preservation of participants' intended meanings and instructional contexts rather than strict word-for-word equivalence.

The corresponding author reviewed the coding framework, participant profiles, selected trajectories, emerging cross-case themes, discrepant cases, and artifact interpretations. The authors clarified code boundaries and resolved interpretive differences through discussion and consensus. Records of coding revisions, theme development, matrix construction, and interpretive decisions were retained as an audit trail. Although these procedures strengthened credibility and consistency, the influence of the facilitator–researcher relationship could not be fully neutralized and remains a limitation of the study.

## 4. Results

The findings are organized according to the two research questions and the analytical phases described in Section 3.4. Section 4.1 presents cross-case patterns in participants' initial AI literacy profiles, whereas Section 4.2 reports within-case trajectories and cross-case patterns across pre-task, interaction, artifact, and post-program data. Participant counts indicate distributions within the eight cases rather than statistical prevalence.

### *4.1. Initial Manifestations of AI Literacy*

The first research question examined participants' initial AI literacy. Their profiles differed in how they understood ChatGPT, anticipated using prompts, framed ethical responsibility, and positioned themselves toward classroom use. Table 3 summarizes the three initial-profile patterns, the participants associated with each pattern, and evidence qualifying the dominant tendencies.

**Table 3.** Cross-case summary of participants' initial AI literacy profiles.

| Integrated Pattern | Cross-Case Distribution | Participants | Qualifying Evidence |
| --- | --- | --- | --- |
| Heterogeneous understandings and prompting readiness | Five participants described learned patterns, prediction, or context-sensitive generation; seven anticipated prompts containing instructional conditions | Generative explanations: T1, T2, T3, T4, T6; contextual prompting: T1–T6, T8 | T5, T7, and T8 relied more on retrieval or information-combination explanations; T7 anticipated broad or template-oriented prompts |
| Verification-centered but uneven ethical awareness | All participants recognized the need to verify AI-generated content; seven raised at least one concern beyond factual accuracy | Verification: T1–T8; broader ethical concerns: T1–T6, T8 | T3 addressed privacy, plagiarism, and attribution most explicitly; T7 focused primarily on reliability and teacher checking |
| Coexisting optimism, caution, and skepticism | One participant showed strong affirmative willingness, six expressed conditional openness, and one maintained predominant skepticism | Affirmative: T7; conditional: T1–T6; skeptical: T8 | T8 remained skeptical about student use but acknowledged limited utility for teacher planning |

#### 4.1.1. Heterogeneous Understandings and Prompting Readiness

As shown in Table 3, five participants described ChatGPT in terms of learned patterns, probabilistic prediction, or context-sensitive generation. T3 provided one of the clearest explanations:

> "[ChatGPT] analyzes the user's question and predicts the most likely phrase from learned data. Although the response may appear creative, it actually operates through patterns and statistics." (T3, pre-task reflection)

The precision of these explanations varied. T4 referred more generally to pattern recognition, whereas T6 combined predictive generation with the inaccurate assumption that responses were produced from data analyzed directly on the web. By contrast, T5, T7, and T8 described ChatGPT mainly through retrieval, recombination, or accumulated interaction data. T7's anticipated use reflected this orientation:

> "For a field I do not know, I would use prompts such as 'Tell me about . . .' or 'Summarize . . ..' I might also provide a lesson topic and lesson-plan template and ask it to 'Complete the following template.'" (T7, pre-task reflection)

Technical understanding and prompting readiness did not align uniformly. Seven participants anticipated specifying instructional conditions such as grade level, lesson topic, learning objective, learner characteristics, classroom context, or desired output. T1, T5, and T8 proposed context-rich prompts despite differing explanations of how ChatGPT operated. All eight expected to modify generated content, but their criteria ranged from general editing to curriculum alignment, learner level, lesson objectives, activity sequence, and feasibility. Initial AI knowledge and prompting readiness therefore appeared in different combinations rather than along a single continuum.

#### 4.1.2. Verification-Centered but Uneven Ethical Awareness

All eight participants recognized that AI-generated content required verification before classroom use. Their concerns most often centered on factual reliability, scientific accuracy, elementary-level appropriateness, and teacher responsibility. T2 connected verification directly to professional judgment:

> "Rather than accepting the information provided by AI unconditionally, it should be verified by comparing it with textbooks or reliable materials. The teacher must evaluate and revise the content based on professional judgment." (T2, pre-task reflection)

T1 likewise proposed comparing outputs with the curriculum, textbooks, and verified external sources, while T4 and T7 emphasized secondary checking through trustworthy information. T5 identified achievement standards, learning objectives, and assessment criteria as bases for deciding whether an output was instructionally appropriate. These responses framed verification as scientific and curricular judgment rather than simple fact-checking.

The scope of ethical concern nevertheless varied. Seven participants raised at least one issue beyond factual accuracy, including bias, privacy, plagiarism, attribution, overreliance, weakened critical thinking, or reduced student agency. T3 articulated the broadest initial scope:

> "I am concerned that children may unconditionally believe what AI says, use another person's work as it is, as in plagiarism, and that personal information may not be protected." (T3, pre-task reflection)

T8 focused more strongly on learning and agency, warning that AI use could become "a 'copy-and-paste' lesson rather than actual learning." Initial ethical awareness was therefore consistently verification-centered but uneven in breadth. Concerns about financial cost, environmental impact, and dependence on a proprietary platform were not substantively evident in the pre-task reflections. However, because the reflection prompts did not explicitly elicit these issues, their absence was not interpreted as evidence that participants lacked awareness of them.

#### 4.1.3. Coexisting Optimism, Caution, and Skepticism

Seven participants expressed at least some willingness to use GenAI in future teaching, although most attached conditions concerning teacher judgment, student thinking, accuracy,

or human interaction. T7 showed the most affirmative position, emphasizing AI's potential for creating experiments, simulations, and visual materials. Most others demonstrated conditional openness. T2 explained:

> "GenAI can be a useful tool for supporting lesson preparation and material development in schools. However, AI should not become the center of education, and we should not forget that interaction between teachers and students is fundamental." (T2, pre-task reflection)

T1, T3, T4, T5, and T6 similarly anticipated benefits such as lesson ideation, alternative explanations, material preparation, and differentiated support, while retaining concerns about misinformation, dependence, creativity, critical thinking, and teacher review.

T8 represented the clearest contrasting profile:

> "I view it negatively. Although AI cannot be excluded from current educational trends, I believe that its negative aspects substantially outweigh its positive ones. Particularly in elementary school, children need to develop the ability to think for themselves." (T8, pre-task reflection)

Even so, T8 proposed using ChatGPT during lesson preparation to anticipate student questions and teacher responses. Initial attitudes therefore ranged from affirmative willingness to predominant skepticism, with conditional openness representing the most common position.

Taken together, participants' initial AI literacy profiles were heterogeneous and internally connected. Differences in technical understanding did not consistently correspond to prompting readiness, ethical scope, or willingness to use GenAI. The first research question was therefore answered through differing combinations of technical understanding, pedagogical readiness, ethical concern, and professional positioning rather than through a uniform low-to-high continuum.

### *4.2. Observable and Context-Specific Changes in AI Literacy*

The second research question examined observable changes across participants' pre-task reflections, ChatGPT interaction logs, reflective journals, instructional artifacts, and final reflective reports. Change was identified only when later statements, interaction practices, revision decisions, or artifacts refined, extended, or differed from earlier evidence. Table 4 summarizes the participant-level trajectories identified through the within-case and cross-case analyses. For each participant, it connects the initial AI use or position with the problem identified, subsequent evaluation or revision, artifact-based evidence, and the resulting analytical interpretation. The following subsections elaborate selected cases that most clearly represent four cross-case patterns: increasingly contextualized and iterative prompting; verification, revision, and tool selection; conditional acceptance; and uneven trajectories involving continuity or awareness–practice inconsistency.

**Table 4.** Participant-level AI literacy trajectories across ChatGPT-Assisted lesson design.

| Participant/ Topic | Initial Position | Problem Identified | Response/Revision | Artifact Outcome | Trajectory Interpretation |
|---|---|---|---|---|---|
| T1: Constellations | Requested a structured lesson specifying grade, unit, duration, and activities | Textbook objectives were incompletely represented, and generated visuals were unsuitable for observation | Added omitted concepts, narrowed the constellation scope, and selected Stellarium for observation | ChatGPT supported lesson organization and creative activities, whereas Stellarium supported actual observation | Refinement of initially structured prompting and function-based tool selection |

**Table 4.** *Cont.*

| Participant/ Topic | Initial Position | Problem Identified | Response/Revision | Artifact Outcome | Trajectory Interpretation |
|---|---|---|---|---|---|
| T2: Seasonal change | Emphasized that AI outputs should be verified using textbooks and reliable sources | In one interaction, the response was judged to contain no factual conflict | Used the response without documented cross-checking or revision in that episode | No corresponding revision was evident, although the final reflection again emphasized verification and teacher responsibility | Episode-specific awareness–practice inconsistency |
| T3: Moon phases | Used ChatGPT for a fictional script and date-specific information about the Moon | The script disclosed too much information, and the Moon response was inaccurate | Revised the script and checked observational data using an external lunar site and Starwalk | Observational evidence became primary, while ChatGPT was used as a secondary checking resource | Enactment of prior ethical awareness and epistemic repositioning |
| T4: Weather | Entered the program with explicit concern about plausible but inaccurate AI responses | Exact-looking weather predictions lacked adequate evidence or explanatory grounds | Questioned the reliability of the predictions and reframed the task around seasonal and regional weather patterns | The final activity avoided unsupported numerical forecasting and emphasized interpretation of characteristic weather conditions | Contextual enactment of a comparatively explicit initial verification orientation |
| T5: Solar system | Requested a Jigsaw lesson structure and recommendations for planet selection | Group movement was impractical, and the rationale for selecting or excluding planets was questionable | Revised the group structure, checked textbooks and external sources, and reconsidered the selected planets | More feasible Jigsaw organization, a create-your-own-planet activity, Padlet sharing, and safety guidance | Classroom-feasibility adjustment and content verification |
| T6: Weather | Initially requested an exact next-day forecast using limited numerical data | Precise values lacked visible reasoning and were insufficiently connected to regional, seasonal, and learner contexts | Reframed prompts around grade, region, season, and learner role and simplified the generated language | The worksheet scaffolded contextual prompting, interpretation, and forecast writing | Scientific and developmental adaptation |
| T7: Moon phases | Initially anticipated broad information-seeking and template-oriented prompts | More specific prompts still produced mismatches between lunar-phase names, appearances, and generated images | Added learner, location, date, and output conditions and simplified image requests | Prompts became more contextualized, but representational inaccuracies remained unresolved | Practical prompting change with attitudinal continuity and persistent output limitations |
| T8: Weather | Initially expressed predominant skepticism because of possible declines in students' thinking and learning | Recognized the usefulness of structured prompts but continued to question reliability, standardized responses, and uneven participation | Added age, role, and purpose to prompts and accepted limited classroom use under teacher control | The final design incorporated contextual prompting while retaining teacher verification and limits on student dependence | Limited practical change and bounded attitudinal acceptance |

#### 4.2.1. Increasingly Contextualized and Iterative Prompting

As summarized in Table 4, six participants showed increasingly contextualized or iterative prompting. This pattern did not necessarily involve a shift from completely general

requests to highly specific ones. Several participants initially included instructional conditions but subsequently refined their prompts after identifying curricular, developmental, or classroom-related limitations. T1 and T5 most clearly demonstrated this evaluative and iterative use of prompting.

T1's initial prompt already specified the grade level, science unit, lesson duration, instructional stages, and proposed activities. After reviewing ChatGPT's response, however, she found that the lesson did not sufficiently reflect the textbook concepts and learning objectives. She consequently added the omitted curricular requirements:

> "I think Activity 1 should include the concepts and content presented in the textbook so that it becomes a more realistic elementary science lesson." (T1, follow-up prompt in the ChatGPT interaction log)

T1 further restricted the lesson to the Big Dipper, Ursa Minor, and Cassiopeia, which were emphasized in the elementary curriculum, and questioned the absence of an activity addressing the meaning of stars and constellations. Her trajectory therefore represented refinement of an already structured prompt through curriculum-based evaluation rather than the acquisition of basic prompting specificity.

T5 also began with a prompt that identified the grade, solar system topic, and Jigsaw model. After evaluating the proposed lesson, she concluded that repeated movement between expert and home groups would be difficult within the available time and might interrupt students' concentration. She retained the initially seated groups as expert groups and used follow-up prompts to develop a student-centered AI activity:

> "I want to add a third activity, creating one's own planet, and I would like students to use GPT in this activity." (T5, follow-up prompt in the ChatGPT interaction log)

The final lesson incorporated a create-your-own-planet activity and used Padlet to share student products without lengthy individual presentations. Iteration was therefore driven by classroom feasibility, student participation, and the instructional purpose assigned to ChatGPT.

Related prompting changes appeared in other trajectories. T6 incorporated the learner's grade, region, season, and roles into weather prompts and transferred this sequence to the worksheet as a scaffold for students. T7 added learner, location, date, target phenomenon, and output conditions, although specific prompts still produced inaccurate Moon images. T3 and T8 also reported that identifying the user, purpose, and instructional context helped produce more relevant responses.

Prompting thus became part of an evaluative and revisable lesson-design process rather than a one-time request. Increased contextual specificity supported curriculum- and learner-sensitive adaptation, but it did not guarantee scientific accuracy, representational reliability, or classroom usability.

#### 4.2.2. From Stated Caution to Verification, Revision, and Tool Selection

Table 4 shows that seven participants demonstrated artifact-linked verification, revision, supplementation, or alternative tool selection. These practices addressed scientific accuracy, curriculum alignment, developmental appropriateness, inquiry orientation, classroom feasibility, and pedagogical coherence. T3's Moon-phase lesson provided the clearest example of how an inaccurate response led to external verification and the reassignment of instructional functions across tools.

T3 initially used ChatGPT to help develop the voice and script of "Dr. Lumen," a fictional character living on the Moon. She retained the narrative structure but revised the script because it disclosed too much information and reduced students' opportunity to investigate. A more consequential problem arose when ChatGPT provided inaccurate information about

the Moon's appearance on a particular date. T3 compared the response with observational information and changed how the tools were used:

> "Because this could also happen when ChatGPT is used in the classroom, we decided to use another site that records the Moon's appearance and to use GPT only to check the answer." (T3, ChatGPT interaction log and open-ended reflection)

In the final instructional artifact, Starwalk and observational data were positioned as the primary sources for examining changes in the Moon's appearance, whereas ChatGPT served as a secondary resource for checking students' inferences. This revision did more than correct an inaccurate response. It reassigned epistemic priority from AI-generated information to observational evidence.

> Related revisions appeared across the other cases. T1 selected Stellarium for constellation observation, while T5 checked ChatGPT's planet selection against textbooks and external sources. Although T4 and T6 independently designed lessons for the same curricular unit and lesson topic, their ChatGPT interactions, generated outputs, revision processes, and individual reflections were distinct and were analyzed separately. In T4's weather-design interaction, ChatGPT was asked to predict the following day's weather from two days of temperature, precipitation, wind-speed, and humidity data for Pyeongchang. It responded, "It is likely to be cloudy in the morning and clear from the afternoon. The high will be 22 °C and the low 16 °C, with 5 mm of precipitation and 65% humidity." T4 noted that the precise figures appeared persuasive but lacked an explanation of how they had been derived from such limited data. The task was consequently reframed around seasonal and regional weather patterns. (T4, ChatGPT interaction log and open-ended reflection)

These trajectories indicate selective acceptance rather than direct adoption. Participants retained ChatGPT where it supported ideation, organization, dialogue, or creativity, but revised, supplemented, rejected, or reassigned its outputs when they did not satisfy scientific or instructional requirements. Verification therefore extended beyond checking whether a statement was factually correct. It also involved examining whether the reasoning was visible, whether representations corresponded to the target phenomenon, whether the content aligned with the curriculum and learner level, and whether another tool was more appropriate for the instructional purpose.

#### 4.2.3. Conditional Acceptance Rather than Uniformly Increased Confidence

The participant-level trajectories in Table 4 indicate that willingness to use GenAI became more differentiated rather than uniformly more positive. Some participants reported increased confidence, whereas others maintained their initial positions while specifying clearer conditions for appropriate use. T5 and T8 represented contrasting forms of conditional acceptance.

T5 began with a generally positive attitude and later reported becoming more positive after experiencing ChatGPT's usefulness for generating ideas and responding flexibly to lesson-design requests. Her increased willingness was nevertheless accompanied by clearer limitations concerning curriculum alignment, misconceptions, privacy, classroom management, and verification:

> "I would avoid presenting what it generates exactly as it is. When used according to the instructional need, it can serve as an excellent assistant." (T5, final reflective report)

Her trajectory therefore reflected greater confidence in using and mediating GenAI, rather than greater trust in every generated response.

T8 showed the clearest movement from predominant skepticism toward bounded acceptance. Before the program, she emphasized possible declines in independent thinking, learning, and creativity. After designing an AI-supported lesson, she acknowledged that limited use could have instructional value:

> "I originally thought that AI would interfere with children's learning. However, while preparing an AI-supported lesson, I came to think that using it to some extent could be acceptable." (T8, final reflective report)

This change did not eliminate her initial concerns. T8 continued to question the reliability of generated information, the possibility of standardized responses, reduced opportunities for students to think independently, and uneven participation when one student controlled an ongoing ChatGPT conversation. Her trajectory therefore represented limited and conditional acceptance rather than general confidence.

Other participants showed related patterns. T1, T3, and T6 reported greater confidence in using GenAI while emphasizing teacher review and adaptation. T2 and T4 largely retained their initially conditional positions. T7 remained positive but developed more concrete caution after encountering inaccurate generated images. Across the cases, increased willingness concerned confidence in professional mediation rather than generalized confidence in the technology.

#### 4.2.4. Uneven Trajectories and Awareness–Practice Gaps

The trajectories summarized in Table 4 also show that the four AI literacy dimensions did not change simultaneously or to the same degree. T2 provided the clearest awareness–practice inconsistency. Before the program, he stated that AI-generated information should be compared with textbooks and reliable sources and evaluated through teacher judgment. In one documented interaction, however, he reported:

> "I used the response as it was because there were no answers that contradicted the facts." (T2, open-ended reflection on the ChatGPT interaction)

No corresponding cross-checking or revision was documented in that episode. His final report again emphasized source verification, privacy, copyright, overreliance, and teacher responsibility. The case therefore did not indicate an absence of ethical awareness; rather, it represented one documented episode in which an articulated verification principle was not enacted in observable practice, rather than evidence of a stable or recurrent pattern.

By contrast, T3 and T4 largely enacted or contextualized principles already evident in their initial profiles. Their trajectories are better interpreted as refinement from comparatively explicit starting positions than as movement from low to high AI literacy. T7 combined more contextualized prompting and clearer recognition of image-generation limitations with a largely stable positive attitude. T8 reported no major change in her overall understanding but identified structured prompting as a practical insight while maintaining substantial skepticism.

Taken together, the second research question revealed context-specific and uneven trajectories rather than a common developmental progression. Participants demonstrated different combinations of newly observable practice, refinement of existing understanding, continuity, bounded attitudinal change, and awareness–practice inconsistency. ChatGPT-assisted lesson design made these differences visible by requiring participants to connect AI use with scientific, curricular, pedagogical, and ethical judgment.

## 5. Discussion

The findings indicate that elementary pre-service teachers' AI literacy was initially heterogeneous and changed through uneven, context-specific trajectories rather than through

a common progression from low to high literacy. Participants entered the program with different combinations of technical understanding, prompting readiness, ethical concern, and professional positioning. During lesson design, observable changes appeared in the contextualization of prompts, evaluation and revision of outputs, selective use of alternative tools, and increasingly conditional views of GenAI. These patterns support multidimensional conceptions of AI literacy in which knowledge, skills, ethics, and attitudes are interrelated but do not necessarily change at the same pace or in the same direction (Ng et al., 2021; Wang et al., 2023; Yue et al., 2026).

The initial profiles also suggest that familiarity with ChatGPT should not be inferred from a single dimension. Some participants articulated pattern-based or probabilistic accounts of response generation, whereas others explained ChatGPT primarily through retrieval or information combination. Participants with less precise technical explanations sometimes anticipated context-rich prompts and pedagogically relevant revisions. This mismatch indicates that conceptual understanding, practical readiness, and professional judgment can appear in different combinations. Previous studies have examined preservice teachers' overall AI literacy and readiness for future classrooms (Ayanwale et al., 2024), their applications and perceptions of GenAI in lesson planning (Hsu et al., 2024), and the relationship among concerns, confidence, and AI literacy (Hur, 2025). The present findings extend this work by showing how these dimensions may diverge and become visible through concrete, subject-specific instructional decisions.

A central finding was that prompting became embedded in an evaluative design cycle. Participants added curriculum boundaries, learner roles, activity purposes, and classroom constraints after examining the responses they received. However, more specific prompts did not consistently yield scientifically accurate, representationally reliable, or classroom-ready outputs. AI literacy became most visible when participants judged whether a response should be retained, revised, supplemented, rejected, or assigned a narrower instructional function. The use of Stellarium or Starwalk for observation, the restructuring of a Jigsaw activity, and the reframing of unsupported weather prediction illustrate function-based tool selection and artifact revision. These findings are consistent with research showing that GenAI can support science lesson planning while requiring verification, contextual adaptation, and professional judgment (Cooper, 2023; Lee & Zhai, 2024; Okulu & Muslu, 2024; Peikos & Stavrou, 2025).

The ethical findings further show that stated awareness and enacted practice should be distinguished. All participants initially recognized some need to verify AI-generated content, and several addressed privacy, plagiarism, attribution, bias, and overreliance. Yet the discrepant T2 case included one documented episode in which an explicit verification principle was not enacted through cross-checking. Ethical AI literacy may therefore require repeated opportunities to connect principles with observable decisions in prompts, lesson materials, and classroom plans. AI literacy has been conceptualized as involving critical engagement with AI systems and consideration of ethical and social implications alongside technical understanding and use (Biagini, 2025; Pinski & Benlian, 2024). In the present study, however, concerns related to access, cost, environmental effects, licensing and ownership of AI-generated materials, transparency about AI use, student-data handling, and dependence on proprietary platforms were not substantively evidenced. Their limited visibility should be interpreted as a limitation of the data elicited by the study instruments rather than as evidence that participants lacked awareness of these issues. These concerns may therefore require explicit attention in future teacher education research and practice.

The artifact analysis also clarifies the scope of the study's claims. Participants' revisions involved scientific accuracy, curriculum alignment, developmental appropriateness, inquiry orientation, classroom feasibility, and pedagogical coherence. These decisions reflected

content-related and pedagogical judgment, but PCK was not independently assessed or used as an analytical framework. The findings should therefore not be interpreted as evidence of PCK development. Likewise, the study did not score lesson quality, compare participants' products, or evaluate classroom effectiveness or student learning. The artifacts served as evidence of how participants translated stated concerns and AI interactions into instructional decisions.

These findings have implications for practice-based teacher education. GenAI preparation should include more than demonstrations of prompt writing or efficient material generation. Pre-service teachers need structured sequences in which they formulate prompts, inspect responses, compare outputs with curricula, textbooks, observational evidence, and reliable external tools, document reasons for revision, and reflect on the ethical consequences of adoption or rejection. Structured ChatGPT-supported critique has been used to support pre-service teachers' lesson-planning work (Cai et al., 2025). Interactive, tool-mediated, and design-based learning activities have also been emphasized as approaches for integrating AI knowledge, skills, ethics, and attitudes through active engagement (Ng et al., 2024; Yue et al., 2026). Consistent with these studies, the present findings suggest that such sequences may make the difference between confidence in using AI and trust in AI-generated content more explicit. Subject-specific lesson design is valuable because it requires participants to coordinate scientific, curricular, pedagogical, and ethical considerations within a concrete task. Teacher education can therefore use ChatGPT-assisted design as a context for examining professional judgment while preserving the teacher's responsibility for deciding what instructional role, if any, should be assigned to GenAI.

## 6. Conclusions

This study examined elementary pre-service teachers' AI literacy in ChatGPT-assisted science lesson design by tracing initial profiles and participant-level trajectories across a ten-week program. Initial AI literacy was heterogeneous rather than uniformly limited: technical understanding, prompting readiness, ethical concern, and willingness to use GenAI appeared in different combinations. During lesson design, participants demonstrated increasingly contextualized prompting, artifact-linked verification and revision, selective use of alternative tools, and predominantly conditional acceptance. Their trajectories also included refinement of explicit initial positions, continuity, limited change, and one episode-specific awareness–practice inconsistency. AI literacy in teacher education should therefore not be interpreted as a single progression from low to high competence.

The study contributes by showing how AI literacy becomes visible through subject-specific instructional decisions. Participants considered whether ChatGPT outputs were scientifically accurate, curriculum-aligned, developmentally appropriate, inquiry-oriented, feasible, and pedagogically coherent. Their engagement involved decisions to retain, revise, supplement, reject, or reassign outputs rather than simply generate materials efficiently. ChatGPT-assisted lesson design can thus serve as a practice-based context for examining professional judgment and responsible AI use, while teacher mediation remains central.

Several limitations define the scope of these conclusions. The study involved eight participants in one Korean teacher education setting and a voluntary, non-credit program facilitated by the first author, whose dual role may have influenced participants' responses. The program used GPT-4o through the web-based ChatGPT interface and involved peer microteaching rather than teaching elementary pupils. PCK was not independently assessed, and the artifact analysis did not score lesson quality or examine classroom effectiveness or student learning. Broader ethical concerns, including access, financial cost, environmental impact, licensing and ownership of AI-generated materials, transparency about AI use, student-data handling, and dependence on proprietary platforms, were not explicitly

elicited by the reflection prompts. Their limited visibility should therefore be interpreted as a limitation of the data-collection instruments rather than as evidence that participants lacked awareness of these issues.

Future research should examine these trajectories in larger and more varied teacher education contexts, across different AI systems and science domains, and during teaching practicum or authentic classroom instruction. Longitudinal studies could also investigate whether stated ethical responsibility is sustained in practice and how subject-specific demands shape teachers' decisions about GenAI use.

**Author Contributions:** Conceptualization, Y.-S.C. and M.-C.K.; methodology, Y.-S.C. and M.-C.K.; validation, Y.-S.C. and M.-C.K.; formal analysis, Y.-S.C. and M.-C.K.; investigation, Y.-S.C.; resources, Y.-S.C. and M.-C.K.; data curation, Y.-S.C.; writing—original draft preparation, Y.-S.C.; writing—review and editing, Y.-S.C. and M.-C.K.; visualization, Y.-S.C.; supervision, M.-C.K.; project administration, M.-C.K.; funding acquisition, M.-C.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a 2-Year Research Grant of Pusan National University.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Cheongju National University of Education (2025CJE-2, with approval granted on 25 April 2025).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The raw data supporting this article will be made available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Ten-Week Instructional Design Program

**Table A1.** Overview of the ten-week instructional design program.

| Week | Main Activities | Objectives |
|---|---|---|
| Week 1 | Introduction to GenAI and pre-task written reflections | Identify participants' initial understanding, skills, ethical concerns, and attitudes toward GenAI |
| Week 2 | Exploration of ChatGPT's educational functions and limitations | Familiarize participants with ChatGPT's possibilities and limitations in educational contexts |
| Week 3 | Review of elementary Earth science curriculum and lesson topics | Connect lesson design tasks with curriculum goals and Earth science content |
| Week 4 | Initial Earth science lesson design using ChatGPT | Generate initial lesson ideas and examine how ChatGPT supports lesson planning |
| Week 5 | Prompt formulation and AI response evaluation | Develop prompting strategies and evaluate AI-generated content for scientific accuracy and appropriateness |
| Week 6 | Lesson plan revision based on AI-generated outputs | Adapt AI-generated content to lesson goals, student level, and curriculum alignment |
| Week 7 | Peer discussion and ethical reflection | Discuss reliability, bias, accuracy, appropriateness, and teacher responsibility |
| Week 8 | Refinement of instructional materials and activities | Improve explanations, inquiry activities, teaching materials, and assessment components |
| Week 9 | Microteaching of AI-supported Earth science lessons | Apply revised lesson plans in a peer-based teaching simulation and reflect on the pedagogical use of ChatGPT |
| Week 10 | Final reflective report and synthesis discussion | Reflect on changes in AI knowledge, AI skills, AI ethics, and AI attitudes |

# Appendix B

*Appendix B.1. Data Sources, Sample Prompts, and Data Collection Instruments*

**Table A2.** Data sources, sample prompts, and analytical focus.

| Data Source | Description | Sample Prompts or Examples | Analytical Focus |
| --- | --- | --- | --- |
| Pre-task written reflections | Open-ended responses completed before the program | "What do you think ChatGPT can and cannot do in science lesson design?" "What concerns do you have about using ChatGPT in elementary science teaching?" | Initial AI knowledge, skills, ethics, and attitudes |
| ChatGPT interaction logs | Participants' prompts, AI-generated responses, and records of revision or rejection | Prompts asking ChatGPT to generate a science lesson idea, revise a scientific explanation, suggest an inquiry activity, or check student-level appropriateness | Prompting strategies, evaluation of AI outputs, instructional decision-making |
| Lesson plans and instructional artifacts | Earth science lesson plans and teaching materials developed during the program | Lesson objectives, instructional procedures, inquiry activities, explanations, assessment ideas, and revised AI-supported materials | Adaptation of AI-generated content and pedagogical reasoning |
| Reflective journals | Written reflections during the program | "How did you evaluate the accuracy and appropriateness of ChatGPT's response?" "What parts of the AI-generated output did you revise, and why?" | Perceived usefulness, limitations, emotional responses, and ethical concerns |
| Observation notes | Researcher notes from program sessions | Notes on participants' discussions, difficulties, prompt revision, peer feedback, and moments of uncertainty or decision-making | Participants' interactions, discussions, difficulties, and reflective moments |
| Final reflective reports | End-of-program reflections structured around AI literacy dimensions | "How has your understanding of ChatGPT changed?" "How have yourprompt and evaluation strategies changed?" "How has your attitude toward using GenAI in science teaching changed?" | Perceived changes in AI knowledge, skills, ethics, and attitudes |

*Appendix B.2. Selected Reflection Prompts*

Representative prompts used before the ten-week instructional design program included the following:

What do you currently understand about how GenAI tools such as ChatGPT work?

If you were to use ChatGPT to design an elementary science lesson, what kinds of prompts would you try, and why?

What concerns do you have about using GenAI tools in elementary science teaching?

How do you feel about using GenAI tools such as ChatGPT in your future teaching practice?

Selected reflective journal prompts included the following:

How did you use ChatGPT during your lesson design process?

How did you evaluate the accuracy and appropriateness of ChatGPT's responses?

What parts of the AI-generated output did you revise, and why?

What difficulties or changes in your thinking did you experience while using ChatGPT?

Representative final reflective report prompts included the following:

How has your understanding of ChatGPT changed?

How have your prompt and evaluation strategies changed?

How has your understanding of teacher responsibility in using AI-generated content changed?

How has your attitude toward using GenAI in science teaching changed?

## References


Anik, M., Raaz, S., & Khan, N. (2025). Embracing AI assistants: Unraveling young researchers' journey with ChatGPT in science education thesis writing. *International Journal of Artificial Intelligence in Education*, *35*(1), 225–244. [CrossRef]

Ayanwale, M., Adelana, O., Bamiro, N., Olatunbosun, S., Idowu, K., & Adewale, K. (2025). Large language models and GenAI in education: Insights from Nigerian in-service teachers through a hybrid ANN-PLS-SEM approach. *F1000Research*, *14*(2), 258. [CrossRef] [PubMed]

Ayanwale, M., Adelana, O., Molefi, R., Adeeko, O., & Ishola, A. (2024). Examining artificial intelligence literacy among pre-service teachers for future classrooms. *Computers and Education Open*, *6*(3), 100179. [CrossRef]

Biagini, G. (2025). Towards an AI-literate future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education*, *35*(1), 2616–2666. [CrossRef]

Biswas, S. (2023). Potential use of chat gpt in global warming. *Annals of Biomedical Engineering*, *51*(6), 1126–1127. [CrossRef] [PubMed]

Brandão, A., Pedro, L., & Zagalo, N. (2024). Teacher professional development for a future with generative artificial intelligence—An integrative literature review. *Digital Education Review*, *45*(2), 151–157. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

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*, 101022. [CrossRef]

Chang, J., Park, J., & Park, J. (2023). Using an artificial intelligence chatbot in scientific inquiry: Focusing on a guided-inquiry activity using inquirybot. *Asia-Pacific Science Education*, *9*(1), 44–74. [CrossRef]

Choi, Y. (2024). A phenomenological approach to the experiences of pre-service earth science teachers utilizing ChatGPT in science instruction. *Journal of the Korean Earth Science Society*, *45*(6), 586–599. [CrossRef]

Choi, Y. (2025a). Exploring the scientific validity of ChatGPT's responses in elementary science for sustainable education. *Sustainability*, *17*(7), 2962. [CrossRef]

Choi, Y. (2025b). Integrating ChatGPT into the design of 5E-based Earth science lessons. *Education Sciences*, *15*(7), 815. [CrossRef]

Choi, Y., & Chung, J. (2025). Analysis of practical characteristics in simulated elementary science lessons earth and space using ChatGPT. *Journal of the Korean Earth Science Society*, *46*(1), 78–95. [CrossRef]

Choi, Y., & Kim, J. (2025). Opportunities and challenges in scientific argumentation using generative artificial intelligence. *Journal of The Korean Association for Science Education*, *45*(1), 61–75. [CrossRef]

Clark, T., Fhaner, M., Stoltzfus, M., & Queen, M. (2024). Using ChatGPT to support lesson planning for the historical experiments of Thomson, Millikan, and Rutherford. *Journal of Chemical Education*, *101*(5), 1992–1999. [CrossRef]

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. [CrossRef]

Creswell, J., & Poth, C. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE publications.

Gatlin, M. (2023). Assessing pre-service teachers' attitudes and perceptions of using artificial intelligence in the classroom. *Texas Educator Preparation*, *7*(2), 1–8. [CrossRef]

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. [CrossRef]

Hsu, H., Mak, J., Werner, J., White-Taylor, J., Geiselhofer, M., Gorman, A., & Capurro, C. (2024). Preliminary study on pre-service teachers' applications and perceptions of generative artificial intelligence for lesson planning. *Journal of Technology and Teacher Education*, *32*(3), 409–437. [CrossRef]

Hur, J. (2025). Fostering AI literacy: Overcoming concerns and nurturing confidence among preservice teachers. *Information and Learning Sciences*, *126*(1/2), 56–74. [CrossRef]

Joksimovic, S., Ifenthaler, D., Marrone, R., De Laat, M., & Siemens, G. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence*, *4*(1), 100138. [CrossRef]

Kotsis, K. (2024). ChatGPT as teacher assistant for physics teaching. *EIKI Journal of Effective Teaching Methods*, *2*(4), 18–27. [CrossRef]

Lee, G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*, *17*, 1643–1660. [CrossRef]

Li, T., Ji, Y., & Zhan, Z. (2024). Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course. *Asia Pacific Journal of Education*, *44*(1), 45–60. [CrossRef]

Ng, D., Leung, J., Chu, S., & Qiao, M. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*(1), 100041. [CrossRef]

Ng, D., Su, J., Leung, J., & Chu, S. (2024). Artificial intelligence (AI) literacy education in secondary schools: A review. *Interactive Learning Environments*, *32*(10), 6204–6224. [CrossRef]

Okulu, H., & Muslu, N. (2024). Designing a course for pre-service science teachers using ChatGPT: What ChatGPT brings to the table. *Interactive Learning Environments*, *32*(10), 7450–7467. [CrossRef]

Peikos, G., & Stavrou, D. (2025). ChatGPT for science lesson planning: An exploratory study based on pedagogical content knowledge. *Education Sciences*, *15*(3), 338. [CrossRef]

Pinski, M., & Benlian, A. (2024). AI literacy for users—A comprehensive review and future research directions of learning methods, components, and effects. *Computers in Human Behavior: Artificial Humans*, *2*(1), 100062. [CrossRef]

Purnama, I., Edi, F., & Wijanarko, T. (2023). ChatGPT for teachers and students in science learning: A systematic literature review. *Jurnal Penelitian Pendidikan IPA*, *9*(10), 760–765. [CrossRef]

Sanusi, I., Ayanwale, M., & Chiu, T. (2024). Investigating the moderating effects of social good and confidence on teachers' intention to prepare school students for artificial intelligence education. *Education and Information Technologies*, *29*(1), 273–295. [CrossRef]

Wang, B., Rau, P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, *42*(9), 1324–1337. [CrossRef]

Yue, M., Jong, M., & Dai, Y. (2022). Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability*, *14*(23), 15620. [CrossRef]

Yue, M., Jong, M., Dai, Y., & Lau, W. (2026). Students as AI literate designers: A pedagogical framework for learning and teaching AI literacy in elementary education. *Journal of Research on Technology in Education*, *58*(3), 593–614. [CrossRef]

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education. *Journal of Science Education and Technology*, *30*(2), 139–149. [CrossRef]

Zhai, X., Nyaaba, M., & Ma, W. (2024). Can generative AI and ChatGPT outperform humans on cognitive-demanding problem-solving tasks in science? *Science & Education*, *34*(1), 640–670. [CrossRef]

Zhang, J., & Zhang, Z. (2024). AI in teacher education: Unlocking new dimensions in teaching support, inclusive learning, and digital literacy. *Journal of Computer Assisted Learning*, *40*(4), 1871–1885. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.