From Prompts to Pedagogy: Preservice Teachers' Experiences with AI-Assisted Math and Science Integrated Lesson Planning

Hoyun Cho
Capital University

Abstract

This study examines the experiences of 34 elementary preservice teachers who completed both traditional and AI-assisted lesson planning for integrated mathematics and science instruction using the 5E instructional model. Through qualitative analysis of structured reflections and AI interaction logs, we found that the process of crafting effective AI prompts served as a powerful catalyst for pedagogical reflection: preservice teachers who struggled to articulate what they wanted from AI were, in effect, confronting gaps in their own mathematical understanding and instructional planning. Three key findings emerged: (a) prompt engineering functions as a form of mathematical pedagogical reasoning, requiring teachers to decompose learning objectives with precision; (b) evaluating AI-generated mathematics content develops critical AI Literacy skills specific to the discipline, with important differences between general-purpose and education-specific platforms; and (c) the structured comparison of traditional and AI-assisted plans creates productive cognitive dissonance that deepens professional judgment.

Keywords: artificial intelligence, technology integration, lesson planning, cross-curricular instruction, teacher preparation, educational technology

1 Introduction

When a preservice teacher types "Write a 5E lesson plan about fractions and science for 3rd grade" into ChatGPT, the AI dutifully produces a multi-page plan covering multiple standards, filled with activities that may or may not be feasible in an actual elementary classroom. The result is often overwhelming and generic. This is not because the AI has failed, but because the prompt itself reveals a fundamental gap: the teacher has not yet clarified which fraction concept, which science standard, what prior knowledge students bring, or what authentic connection between the two disciplines would serve students best.

This gap between a vague instructional intention and the specificity required for effective teaching is not new. What is new is that AI tools make this gap immediately visible in ways that traditional planning does not. When a teacher plans a lesson using textbooks and curriculum guides, vague thinking can remain hidden beneath familiar routines. When that same teacher must articulate their instructional goals precisely enough for an AI to produce useful output, the imprecision becomes impossible to ignore.

This article reports on a study that began as an investigation of preservice teachers' experiences with AI-assisted lesson planning but revealed something unexpected: the process of learning to write effective AI prompts functioned as a powerful mechanism for developing mathematical pedagogical reasoning. Preservice teachers who struggled to prompt AI effectively were not simply lacking technical skills. They were discovering the contours of their own mathematical understanding and instructional thinking. We write this article for mathematics teacher educators and methods course instructors who are navigating the practical question of whether and how to incorporate AI tools into their courses.

2 Literature Review

The integration of AI into mathematics teacher preparation sits at the intersection of several established and emerging lines of research. We briefly review three areas most directly relevant to this study: AI as an educational technology tool, the concept of Critical AI Literacy, and the role of prompt engineering in pedagogical contexts.

2.1 AI Tools in Mathematics Education

The emergence of generative AI represents a distinct paradigm for technology integration in education. Unlike previous educational technologies that primarily enhanced existing instructional practices, AI systems function as generative partners that can produce sophisticated pedagogical content (Kasneci et al., 2023; Mollick & Mollick, 2023). Mollick (2024) describes navigating a "Jagged Frontier," where AI may excel at certain complex tasks, such as generating diverse activity ideas and differentiation strategies, while failing at others. In mathematics and science education, the challenge of AI "hallucinations," where AI produces plausible but incorrect information, remains significant. ChatGPT has been documented to frequently produce unreliable responses containing conceptual errors in STEM contexts (Gregorcic & Pendrill, 2023; Kortemeyer, 2023). More recently, Walkington (2025) found that AI tools tend to produce surface-level rather than authentic connections when generating cross-curricular mathematical tasks, and Malik et al. (2025) documented AI's limitations in generating accurate visual representations essential to mathematics instruction. These findings underscore that AI integration in mathematics education requires teachers to possess robust content knowledge not only to plan instruction but also to critically evaluate machine-generated alternatives.

2.2 Critical AI Literacy

The capacity to evaluate AI outputs for accuracy, bias, and pedagogical appropriateness has been termed Critical AI Literacy (Ng et al., 2021; Laupichler et al., 2023). This concept extends traditional notions of digital literacy by emphasizing that effective use of AI requires more than technical proficiency, and it demands deep subject-matter expertise and professional judgment. In teacher education specifically, Celik (2023) proposed the Intelligent-TPACK framework, arguing that AI integration demands AI-specific technological and pedagogical knowledge with particular attention to ethical dimensions. For mathematics teacher educators, Critical AI Literacy is especially consequential because the precision demanded by mathematical content makes evaluation of AI-generated material both more important and more cognitively demanding than in many other disciplines.

2.3 Prompt Engineering as Pedagogical Practice

An emerging line of inquiry examines how the process of crafting prompts for AI systems may itself serve as a form of pedagogical reasoning. When educators must articulate their instructional goals with sufficient precision for AI to produce useful output, they engage in a process of decomposition and clarification that mirrors backward design thinking (Trust et al., 2023). Mishra et al. (2023) argue that TPACK must be reconceptualized for the age of generative AI, with prompt engineering representing a new form of technological knowledge deeply intertwined with pedagogical reasoning. This perspective suggests that the struggle to write effective prompts is not merely a technical challenge but also an occasion for professional reflection. This hypothesis is examined empirically in this study.

3 Method

This article draws on the same dataset as a companion study (Cho, 2026) that examines broader theoretical implications for teacher education through the lenses of TPACK, the Technology Acceptance Model, and Activity Theory. This article focuses on mathematics-related findings with direct implications for practice.

3.1 Participants and Setting

The study took place in an Integrated Mathematics and Science Pedagogy course for elementary (K–5) preservice teachers at a Midwestern university. Thirty-four participants across two consecutive semesters enrolled in the study. All participants had previously completed a curriculum development course covering lesson, unit, and year planning, as well as mathematics and science content courses and an educational technology course. Of the 34 participants, 47% had minimal or no prior AI experience, while 53% had moderate familiarity (primarily with ChatGPT). Pseudonyms are used throughout to protect participant identities.

3.2 Study Design

Participants completed a structured six-week sequence consisting of four phases: Traditional Planning (2 weeks), AI Training (1 week), AI-Assisted Planning (2 weeks), and Comparative Reflection (1 week). During the traditional planning phase, participants developed integrated mathematics and science lesson plans using the 5E instructional model (Engage, Explore, Explain, Elaborate, Evaluate) with conventional resources such as textbooks, curriculum guides, and online databases. During the AI training phase, participants received explicit instruction in prompt engineering, AI literacy, and output evaluation criteria. During the AI-assisted planning phase, participants developed parallel lesson plans on the same topic using AI tools, including ChatGPT (a general-purpose AI system), and EduAide.ai and Magic School AI (education-specific platforms). During the final phase, structured reflection prompts guided participants to compare their two planning experiences across specific dimensions.

Several design features are important for understanding the findings. First, all participants completed traditional planning before AI-assisted planning, establishing a baseline against which they could evaluate AI's contributions. Second, participants received one week of explicit instruction in prompt engineering before using AI tools for planning. Third, participants were exposed to both general-purpose AI systems and education-specific platforms, enabling comparison across tool types. Fourth, structured reflection prompts guided participants to compare their two planning experiences across specific dimensions: creativity, cross-curricular integration quality, differentiation, and professional growth.

3.3 Data Sources and Analysis

Data sources included structured reflections addressing four dimensions (planning process, integration quality, student learning, and professional growth) and AI interaction logs documenting prompt-response sequences. Data were analyzed using Braun and Clarke's (2006) six-phase thematic analysis approach. Both confirming and disconfirming evidence were included for each theme to maintain analytical balance. Participant quotes were preserved verbatim to maintain authenticity. The author served as the course instructor, and several measures were taken to mitigate social desirability bias: reflections were not graded based on attitudes toward AI, prompts were designed to elicit both positive and negative experiences, and disconfirming evidence was deliberately sought during analysis.

4 Findings

Three interrelated findings emerged from the analysis. Each addresses a distinct dimension of how AI-assisted lesson planning intersects with mathematical thinking and professional development.

4.1 Finding 1: Prompt Engineering as Mathematical Pedagogical Reasoning

The most significant finding of this study was not about AI at all, but about how prompting AI forced preservice teachers to think more precisely about mathematics instruction. Participants across both cohorts described a consistent pattern: initial prompts produced unsatisfying results, which led to iterative refinement, which in turn required increasingly specific mathematical and pedagogical thinking.

4.1.1 From Vague Intentions to Precise Objectives

Rachel, a participant who entered the study with strong resistance to AI, described her early experience candidly:

> I was always very against AI. As we were working on the project, I started to see how much time and effort it can save us. But at first, our prompts were too broad, resulting in entire lesson plans instead of the specific section we needed. Once we figured out the right wording, it became easier.

What Rachel describes as "figuring out the right wording" was, in pedagogical terms, learning to decompose a complex instructional goal into its constituent elements: identifying the specific standard, the particular phase of instruction, the target student population, and the appropriate level of complexity. Her journey from resistance through frustration to pragmatic appreciation was representative of a common pattern, though not universal.

Another participant articulated this connection explicitly: "The AI was strongest when I was very specific in my prompts." When asked what "specific" meant in practice, participants described a process remarkably similar to backward design thinking: starting with a clear learning objective, identifying the evidence that would demonstrate understanding, and only then designing the learning experience. The difference was that AI made the cost of imprecision immediate and concrete. A vague prompt produced a vague lesson plan, providing instant feedback on the teacher's clarity of thought.

Table 1 shows how prompt refinement parallels the development of mathematical pedagogical reasoning. The examples are drawn from participants' actual AI interaction logs and demonstrate the progression from ineffective to effective prompts, mirroring the growth in specificity and mathematical precision that teacher educators work to develop throughout preparation programs.

Table 1. *Prompt Engineering Progression: From Vague to Mathematically Precise*

Stage	Example Prompt	Typical AI Output	Mathematical Thinking Revealed
Stage 1: Unfocused	“Write a 5E lesson plan about fractions and science for 3rd grade.”	A complete but generic multi-page plan covering multiple standards without clear focus	Teacher has not yet identified which fraction concept, which science connection, or what students need to learn
Stage 2: Partially focused	“Create an Explore activity for 3rd-grade fractions and measurement.”	A single activity, but may not align with specific standards or student needs	Teacher has identified a 5E phase and general topics but hasn’t specified standards, time constraints, or learner context
Stage 3: Mathematically precise	“Generate three Explore activities integrating Ohio Math Standard 3.NF.1 with Ohio Science Standard 3.ESS2.1. Activities should involve hands-on data collection and take no more than 20 minutes each.”	Focused, standards-aligned activities with realistic time frames	Teacher has decomposed the lesson into specific standards, identified the conceptual connection, specified pedagogical approach, and set practical constraints
Stage 4: Iterative refinement	“The activity you suggested would take too long for a 10-minute opener. Revise it to a 5-minute warm-up using a real-world scenario about measuring rainfall.”	Revised activity with appropriate scope and real-world context	Teacher is evaluating AI output against practical knowledge, content knowledge, and pedagogical knowledge

4.1.2 Prompt Engineering as Metacognitive Practice

Beyond the technical dimension, participants described prompt engineering as a metacognitive experience that revealed the quality of their own instructional thinking. One participant captured this dimension:

> I found myself having to re-write prompts to be more specific, or word them differently so the AI tool could understand what we needed for our lesson. Every time I rewrote, I realized I was getting clearer about what I actually wanted students to learn.

While this participant framed the challenge as communicating with AI, the deeper process was one of clarifying her own instructional thinking. Another reflected on the dual competencies that developed simultaneously: “I learned how to write effective AI prompts that yield usable ideas. I improved my ability to sort through recommendations to determine what fits my students.” This insight highlights the development of prompt engineering (a technical skill) and pedagogical evaluation (a professional skill) occurring together.

A third participant articulated the novelty of this skill set: “My newest skill is learning how to integrate AI into traditional lesson planning and making it work out for me.” Yet not all participants experienced this metacognitive dimension positively. One expressed frustration: “AI felt like cheating. It put me in a mindset where, instead of creating my own ideas, I was wrestling with someone else’s.” This resistance, rather than indicating failure, reveals an important tension between efficiency and professional ownership that teacher educators should anticipate and address.

For mathematics teacher educators, this finding has immediate practical implications. The iterative prompt–evaluate–refine cycle mirrors the mathematical reasoning process we want preservice teachers to develop: specify a problem clearly, evaluate the result against criteria, and revise the approach based on feedback. In this sense, prompt engineering is not a distraction from mathematical thinking but a new context in which mathematical thinking can be practiced and made visible.

4.2 Finding 2: Evaluating AI-Generated Mathematics Content

If prompt engineering develops the skill of articulating mathematical goals, evaluating AI output develops the complementary skill of assessing mathematical quality, a capacity we term mathematics-specific Critical AI Literacy. This finding addresses both the substance of what participants discovered when evaluating AI-generated content and the important differences between AI platform types.

4.2.1 What AI Gets Wrong in Mathematics

Participants identified several systematic patterns in AI-generated mathematics content that required correction. These concerns clustered around four areas: timing feasibility, grade-level appropriateness, integration balance, and differentiation adequacy.

Regarding timing, one participant described:

> In the Explaining phase, some ideas went too deep and would take almost 45 minutes to cover what should have been addressed in 30 minutes. The time on the lesson might not have been realistic for an elementary classroom. It takes a long time to get all the kids to do what they're supposed to be doing.

This observation reflects a form of practical knowledge, understanding the pace and routines of elementary classrooms, that AI does not possess. Another participant noted integration imbalances: "AI could give more information on math, rather than the science portion or vice versa, but struggled to balance both simultaneously." A third identified the gap between AI-generated suggestions and classroom realities: "AI gives more of idealistic or imaginative ideas that may not translate well to actual classroom implementation."

Perhaps most importantly, participants discovered that AI's differentiation suggestions frequently fell short of what mathematics teaching requires. One explained:

> One challenge was that AI rarely included modifications we need to incorporate daily. While I managed to get this information, it took considerable time changing my input. AI tools might not take into account the needs of ELLs or students with specific learning disabilities.

Another participant contrasted AI-generated differentiation with teacher-generated approaches: "Lessons created by the teacher will naturally include differentiation intertwined for all learners. The teacher has first-hand knowledge of what modifications are needed based on year-long interactions with the class." A third noted the irreplaceable nature of contextual knowledge: "AI does not know my students as well as I do. It can't detect their moods, personalities, or dynamics within the classroom." These observations highlight that differentiating mathematics instruction for diverse learners requires contextual knowledge that AI simply does not possess.

4.2.2 Platform Differences: General-Purpose vs. Education-Specific AI

Participants who used multiple platforms observed that the nature of strengths, limitations, and required corrections varied systematically depending on the type of AI system. These differences have direct implications for how mathematics teacher educators guide tool selection.

Education-specific platforms (EduAide.ai, Magic School AI) provided built-in structural scaffolding aligned with instructional frameworks. One participant explained: "Using Eduaid.ai, the AI provided a complete 5E lesson structure, making it easier to visualize how the lesson would flow. I didn't feel stuck." Another noted: "Magic School in particular has tools for a wide range of educational activities, making the planning process more streamlined." These platforms also embedded standards alignment as a default feature and included differentiation as a standard component. A third participant noted the practical advantage: "The AI program gives you the lesson and what materials you need. It creates a worksheet for students with modified work already created."

General-purpose systems (ChatGPT), by contrast, required more detailed prompting to produce organized structures but offered a broader range of creativity. One participant described how ChatGPT "suggested ideas I wouldn't have thought of, like digital moon simulations"—resources that illustrate AI's capacity to identify novel mathematical connections. Another valued ChatGPT's flexibility: "ChatGPT helped rapidly generate options, expand my examples, and offer alternative phrasings when I was stuck." However, some participants valued familiarity over novelty: "ChatGPT was the one me and my group were familiar with, so we could navigate it more easily despite needing more specific prompts."

These systematic differences led many participants to develop hybrid strategies. Maria described a sequential approach: "I started with ChatGPT to generate ideas, then used Eduaid to organize them into a real lesson." This adaptive behavior reflects sophisticated technological pedagogical knowledge, as participants learned to match specific platform affordances to specific planning tasks rather than relying on any single tool. Table 2 summarizes the key differences participants reported.

Table 2. *Participant-Reported Differences Between AI Platform Types*

Dimension	General-Purpose (ChatGPT)	Education-Specific*
Structural scaffolding	Required detailed prompting; participants often received entire lesson plans rather than targeted sections	Provided built-in frameworks aligned with instructional models (e.g., 5E structure)
Creative range	Generated unexpected, novel ideas (e.g., digital simulations, match-up games)	Generated curriculum-aligned content but perceived as more formulaic
Standards alignment	Required explicit prompting and manual verification	Embedded standards alignment as a default feature
Learning curve	Higher initial barrier due to prompt engineering demands	Lower barrier through structured interfaces
Differentiation	Generated strategies when explicitly prompted; broader range but less structured	Included differentiation as standard; more systematic but sometimes generic
Optimal use case	Brainstorming, creative ideation, exploring novel connections	Structured lesson development, standards-aligned planning, ready-to-use materials

Note. *Education-specific platforms include EduAide.ai and Magic School AI. Differences are based on participant reflections and may not reflect current platform capabilities, which evolve rapidly.

4.2.3 From Passive Consumers to Critical Evaluators

The structured comparison between traditional and AI-assisted planning was crucial for developing evaluative skills. Participants who had already grappled with the challenges of cross-curricular integration in their traditional plans were better positioned to recognize when AI-generated integration was superficial. One explained:

> The traditional method made the math and science connection feel very natural. Because I was thinking through everything myself, linking fractions to moon phases made sense. It felt more intentional.

Another echoed: "Traditional approaches foster more authenticity because every step is taken on purpose." In contrast, several observed that AI tended to treat subjects sequentially: "The AI-assisted plan seemed to keep subjects separate. It had me focusing on science until the final step, where we added the math portion." Another noted: "AI suggestions treated both content areas equally in some cases—math operations were woven directly into scientific measurement tasks—but this was inconsistent." These observations align with Walkington's (2025) finding that AI tools tend to produce surface-level rather than authentic cross-curricular connections.

However, opinions were not unanimous. Some participants found AI facilitated more authentic connections: "The AI approach led to more authentic connections because it was able to come up with so many different discussion topics that I had not thought of." Another described AI identifying unexpected cross-disciplinary links: "AI gave me new examples where subjects overlapped authentically—predicting plant growth using line-plot patterns, for instance." This variation suggests that integration quality depends on both the specific prompts used and the teacher's skill in evaluating and adapting AI suggestions, reinforcing the centrality of Critical AI Literacy.

Based on participants' experiences, Table 3 presents a mathematics-specific quality assurance checklist for evaluating AI-generated lesson content. We have found this checklist useful for structuring peer review sessions in our methods courses.

Table 3. *Mathematics-Specific Quality Assurance Checklist for AI-Generated Content*

Criterion	Guiding Questions	Common AI Pitfall
Standards alignment	Does the content address the specified Ohio Learning Standards? Are both mathematics and science standards treated with equal depth?	AI often emphasizes one subject over the other, creating integration that feels like "adding math at the end"
Conceptual accuracy	Are mathematical representations correct and grade-appropriate? Are science concepts accurately connected?	AI may generate plausible but incorrect examples, particularly for fraction models and measurement conversions
Time feasibility	Is the activity realistic for the allocated time? Does it account for transitions and elementary pacing?	AI routinely underestimates time: "Some ideas would take 45 minutes for what should have been 30"
Integration authenticity	Do mathematics and science connect meaningfully throughout the lesson, or is one an add-on?	AI may produce parallel activities rather than genuinely integrated experiences
Mathematical depth	Does the lesson develop conceptual understanding, not just procedural fluency?	AI may default to procedural tasks unless specifically prompted for conceptual depth
Differentiation adequacy	Are specific accommodations included for ELLs, students with IEPs, advanced learners, and struggling learners?	AI produces generic suggestions rather than specific strategies (e.g., "pre-made graph templates for students still developing data skills")
Practical feasibility	Are suggested materials and technologies available in a typical elementary setting?	AI generates "idealistic or imaginative ideas that may not translate well to actual classroom implementation"

4.3 Finding 3: The Productive Tension Between Traditional and AI-Assisted Planning

The most nuanced insight from this study emerged from participants' reflections on completing both planning approaches in sequence. Rather than viewing AI as simply better or worse than traditional methods, most participants arrived at a position of productive tension, recognizing that each approach developed different professional capacities.

4.3.1 What Traditional Planning Teaches That AI Cannot

Several participants articulated the irreplaceable value of the traditional planning experience. Emily, who consistently advocated for the authenticity of traditional integration, elaborated:

> I felt like when using AI it completely removed that creative component and I can see how a teacher might feel disconnected to the lesson. When I planned traditionally, every choice felt like mine. With AI, I was reacting to someone else's ideas instead of building my own.

Emily's observation reveals something beyond a preference for traditional methods: it articulates a relationship between creative ownership and instructional quality that has important implications for how we introduce AI tools to preservice teachers. The cognitive labor of traditional planning, such as the hours spent searching for connections, the struggle to align standards, and the creative act of designing activities from scratch, develops a form of professional ownership that AI-assisted planning does not replicate.

David, an older student with no prior AI experience, remained cautious throughout the study:

> I actually spent less time on the traditional lesson plan than with AI. Maybe that's because I belong to an older generation. The AI was an extra thing I had to figure out on top of the actual teaching part. I kept going back and forth between what it gave me and what I knew would work.

David's experience is a valuable reminder that AI tools do not automatically reduce planning time, particularly for those who must simultaneously learn the technology and the pedagogy. His perspective challenges the assumption that AI is always more efficient and highlights the importance of considering prior technology experience when designing AI-integrated coursework.

4.3.2 What AI-Assisted Planning Adds

At the same time, participants identified genuine contributions that AI made to their mathematical thinking. Jessica described the shift in cognitive approach:

> When I completed the 5E lesson plan traditionally, it took me several hours and a few days to organize the information. However, when I used AI, it only took about 20 seconds with clearer explanations. When I turned to AI, the process felt more like brainstorming with an instant colleague.

Jessica's metaphor of AI as "colleague" signals a reconceptualization of lesson planning from individual authorship to collaborative curation. The dramatic reduction in initial generation time freed her to redirect cognitive effort toward evaluation and refinement. This higher-order thinking may be more professionally valuable than content generation. Another participant echoed: "AI generates options quickly, giving me more time to tailor the lesson to my students' needs."

AI's creative contributions also expanded mathematical possibilities. One described how AI generated "digital moon simulations and various matchup games for fractions." Maria captured this expansive quality: "The AI lesson allowed me to take my lesson a step further. The AI lesson expanded the horizon of what I thought the lesson could be." Importantly, this sense of expanded possibility did not diminish Maria's professional agency; rather, it repositioned her role from sole generator of ideas to curator and adapter of a broader set of options.

However, AI's differentiation capabilities also generated surprise. One participant reflected: "One unexpected insight was AI's extensive knowledge of differentiation. I was taken aback by how thoroughly it addressed ways to work with students at different proficiency levels." Another described specific AI suggestions: "AI suggested tiered support pre-made graph templates for struggling students and open-ended digital tools for high-ability students." These examples demonstrate that AI can serve as a resource for expanding preservice teachers' repertoire of differentiation strategies, even as it requires human judgment to evaluate their feasibility.

4.3.3 Toward a Both/And Pedagogy

By the study's conclusion, most participants had moved beyond an either/or framing to articulate a both/and position. One summarized: "Using both methods in the classroom could be the most effective approach, with AI serving as a helpful partner, not the driver, but a tool I can pull ideas from." Another articulated a specific vision: "AI will serve as a rapid idea generator, a draft maker for student handouts, and a differentiation assistant, while I maintain responsibility for final instructional decisions."

Rachel's evolution exemplifies this trajectory. Having begun as one of the study's most resistant participants, she concluded: "AI will never replace my expertise as an educator. Instead, it will serve as a tool that enhances my planning efficiency." Rachel's statement reflects a sophisticated understanding of professional expertise, in which the teacher's value lies not in generating content from scratch but in making informed judgments about how to contextualize, adapt, and deliver instruction.

Maria articulated this emerging professional identity most explicitly: "My role will be to curate, adapt, and align AI outputs with my teaching philosophy and the unique needs of my students." This shift from "content deliverer" to "learning facilitator" represents a reconceptualization of professional identity that has significant implications for teacher preparation.

5 Discussion

The three findings presented above converge on a central insight: when structured thoughtfully, AI-assisted lesson planning functions less as a shortcut to instruction and more as a catalyst for professional reasoning. In this section, we situate these findings within the broader literature and discuss their implications for mathematics teacher education.

5.1 Prompt Engineering and Mathematical Reasoning

Finding 1 suggests that the iterative prompt–evaluate–refine cycle functions as a new form of mathematical pedagogical reasoning. This finding extends Mishra et al.'s (2023) argument that TPACK must be reconceptualized for generative AI by providing empirical evidence that the Technological Knowledge component must expand to encompass prompt engineering, not as a standalone technical skill but as a competency deeply intertwined with Pedagogical Content Knowledge. When participants discovered that "the AI was strongest when I was very specific in my prompts," they were articulating a relationship between prompt precision and mathematical clarity that has not been documented in prior research on AI in teacher education.

This finding also resonates with Celik's (2023) Intelligent-TPACK framework, which posits that AI integration demands AI-specific technological and pedagogical knowledge. Our data suggest that this integration is not merely additive but transformative. The process of learning to prompt AI effectively changed how participants thought about lesson planning itself, making explicit the decomposition of learning objectives that often remains implicit in traditional planning.

5.2 Critical AI Literacy in Mathematics Contexts

Finding 2 contributes to the growing literature on Critical AI Literacy (Ng et al., 2021; Laupichler et al., 2023) by documenting how this capacity manifests in a specific professional domain. The quality assurance challenges participants identified, including timing feasibility, differentiation inadequacy, and integration superficiality, represent domain-specific forms of AI limitation that differ qualitatively from the factual errors typically examined in AI accuracy studies (Gregorcic & Pendrill, 2023; Kortemeyer, 2023). A 45-minute activity scheduled for a 30-minute block is not a "hallucination" in the traditional sense but a failure of practical pedagogical knowledge. Identifying such failures requires the kind of contextual expertise that only experienced educators possess.

The platform comparison data extend the Technology Acceptance Model's construct of perceived ease of use by showing that "ease" varies systematically by planning task. Education-specific platforms lower the barrier for structured lesson development, while general-purpose systems better support creative ideation. Participants' organic development of hybrid strategies, using ChatGPT for brainstorming and education-specific tools for structured planning, demonstrates a form of strategic technological knowledge that teacher preparation programs should cultivate deliberately.

5.3 The Pedagogical Value of Productive Tension

Finding 3 reveals that the structured comparison between traditional and AI-assisted planning generates what might be called productive cognitive dissonance. Participants who experienced both approaches were better equipped to articulate what each contributed to their professional development and what each lacked. This finding aligns with Activity Theory's concept of secondary contradictions (Engeström, 2001): the introduction of AI as a mediating tool created tensions between efficiency and authenticity, between AI's generative capacity and teachers' professional norms regarding creative ownership, that ultimately catalyzed new forms of professional practice.

Crucially, these contradictions functioned not as barriers to adoption but as catalysts for professional growth. The "both/and" position most participants reached by the study's conclusion resolves these tensions through what Engeström terms expansive learning, the development of new professional practices that transcend the limitations of either approach alone.

6 Implications for Mathematics Teacher Educators

Based on our findings, we offer four practical recommendations for mathematics methods course instructors considering AI integration.

6.1 Start With Traditional Planning

Our design, with traditional planning first and AI-assisted planning second, was pedagogically essential, not merely methodological. Participants who had wrestled with cross-curricular integration independently were better positioned to evaluate AI's integration attempts. We recommend that instructors resist the temptation to introduce AI tools early in the semester. The cognitive labor of traditional planning builds the very pedagogical content knowledge that makes AI tools useful. Without that foundation, AI becomes a crutch rather than a catalyst.

6.2 Teach Prompt Engineering as Mathematical Thinking

Rather than treating prompt engineering as a technical skill, we recommend framing it as an extension of mathematical pedagogical reasoning. When a preservice teacher writes a vague prompt, the instructor's response should not be "You need a better prompt" but rather "What specific mathematical idea do you want students to understand? What would that understanding look like? What activity would help them get there?" The prompt engineering process then becomes a scaffold for backward design thinking. Table 1 can be used as a discussion tool to help preservice teachers recognize that effective prompting is inseparable from clear mathematical thinking.

6.3 Use the Quality Assurance Checklist as a Teaching Tool

Table 3's checklist was developed from participants' actual experiences and reflects the specific ways AI-generated mathematics content tends to fall short. We recommend using this checklist during peer review sessions where preservice teachers evaluate each other's AI-assisted lesson plans. The checklist makes evaluation criteria explicit and provides a common vocabulary for discussing mathematical quality. Over time, the goal is for preservice teachers to internalize these criteria, at which point they have developed the Critical AI Literacy that effective mathematics teaching in an AI-mediated world requires.

6.4 Structure Comparative Reflection

The six-week sequence we used (Table 4) can be adapted to various course structures. The key design principle is comparison: preservice teachers must experience both traditional and AI-assisted planning to develop the evaluative judgment that neither approach alone provides. We recommend ending the sequence with a structured reflection session in which participants articulate their own “both/and” position, not as a forced conclusion but as a genuine professional stance informed by direct experience.

Table 4. *Suggested Implementation Timeline for Methods Courses*

Week	Focus	Key Activities
1–2	Traditional planning foundation	Preservice teachers develop integrated math-science lesson plans using traditional methods. This establishes the baseline for comparison and builds the pedagogical content knowledge needed to evaluate AI output.
3	AI literacy and prompt engineering	Introduction to AI tools; guided prompt engineering exercises using examples from Table 1; discussion of effective prompts; explicit connection between prompt precision and mathematical thinking clarity.
4–5	AI-assisted planning with scaffolding	Preservice teachers develop parallel lesson plans using AI tools. Apply quality assurance checklist (Table 3). Document prompt iterations. Require written justification for all adaptations made to AI output.
6	Comparative reflection	Structured reflection comparing both experiences; peer review of plans; facilitated discussion of professional implications; articulation of personal “both/and” positions on AI’s role in practice.

7 Limitations and Future Directions

Several limitations should be acknowledged. The data were drawn from a single institution, limiting generalizability. All participants were elementary (K–5) preservice teachers; findings may differ for secondary preservice teachers who possess deeper disciplinary specialization. The study focused specifically on integrated mathematics-science planning using the 5E framework, and results may not transfer directly to other content areas or instructional models. The author served as the course instructor, introducing the potential for social desirability bias despite the mitigation measures described in the Method section.

The rapid evolution of AI capabilities means that specific platform comparisons may become outdated; however, the broader patterns of human-AI interaction documented here, including the importance of prompt specificity, the need for quality assurance, and the development of professional judgment through comparative experience, are likely to remain relevant as specific tools evolve.

This study captured initial experiences within a structured course context; longitudinal research is needed to examine how AI integration practices develop as preservice teachers transition into student teaching and in-service practice. The study did not assess the effectiveness of AI-assisted lesson plans in actual classroom implementation, which represents an important next step. Future research should also examine whether the patterns identified here hold across different teacher preparation contexts, content areas, and grade levels. Comparative studies examining how varying levels of AI training and scaffolding affect integration outcomes would be particularly valuable for informing program design.

8 Concluding Thoughts

The preservice teachers in this study taught us something that the broader AI-in-education discourse often overlooks: the most valuable aspect of AI-assisted lesson planning may not be the lesson plan it produces, but the thinking process it demands. When Rachel struggled to write a prompt that would generate a useful Explore activity, she was not failing at technology; she was succeeding at the hard work of clarifying her mathematical intentions.

When Emily felt "disconnected" from her AI-assisted lesson, she was articulating a profound insight about the relationship between professional ownership and instructional quality. When David found traditional planning faster than AI-assisted planning, he was demonstrating that expertise cannot be shortcut.

These are not stories of technology triumph or failure. They are stories of preservice teachers developing professional judgment and learning to think mathematically, pedagogically, and critically about tools that will shape their careers. As mathematics teacher educators, our task is not to teach preservice teachers to use AI. It is to ensure that the process of engaging with AI deepens their understanding of what it means to teach mathematics well.

AI Use Disclosure

AI was used to assist with improving the clarity, coherence, and grammatical accuracy of the English-language text of this manuscript. AI did not contribute to research design, data analysis, interpretation of results, or the generation of original content. All intellectual contributions, analytical decisions, and conclusions are the authors' own.

Appendix: Implementation Guide for Teacher Educators

This appendix provides practical guidance for teacher educators who wish to integrate AI-assisted lesson planning into their courses. Recommendations are drawn from findings of the present study and organized around common implementation considerations.

A. Common Challenges and Recommended Responses

Table A1 presents challenges frequently encountered during AI-assisted lesson planning instruction, with evidence-based recommendations for addressing each.

Table 5. *Anticipated Challenges in AI-Assisted Lesson Planning Instruction*

Challenge	Description	Recommended Response
"AI is cheating" resistance	Some preservice teachers may view AI use as undermining professional skill development.	Frame AI as a collaborative partner. Use structured comparison tasks to let participants experience both approaches. Emphasize that evaluating AI output requires significant expertise.
Over-reliance on AI output	Some participants may accept AI-generated content without critical evaluation.	Require explicit documentation of modifications and justifications. Use peer review of AI-assisted plans to develop evaluation skills.
Prompt engineering frustration	Initial prompts often produce generic or overwhelming output, leading to frustration.	Begin with structured prompt templates (Table 1), then gradually move toward open-ended prompting. Model the iterative refinement process.
Uneven platform experience	Participants may have different experiences depending on their chosen platform.	Expose participants to multiple platforms. Discuss platform differences explicitly (Table 2) and guide development of hybrid strategies.
Inadequate differentiation	AI-generated accommodations may be generic and not tailored to specific student populations.	Require participants to apply AI suggestions to case-study students with defined characteristics. Evaluate whether AI accommodations would actually be effective.

B. Sample Reflection Prompts

The following reflection prompts were used to guide participants' comparative analysis. They can be adapted for use in methods courses.

Planning Process: How did each approach (traditional and AI-assisted) influence your creativity in lesson design? What unexpected challenges or insights emerged during each planning method?

Integration Quality: Which approach led to more authentic connections between mathematics and science? What unique integration opportunities emerged from each approach?

Student Learning: How might each approach affect different types of learners? What scaffolding strategies emerged from each planning method?

Professional Growth: What new skills have you developed through this project? What role do you see AI playing in your future planning process?

References

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

Cho, H. (2026). Transforming preservice teachers' lesson planning in integrated mathematics and science through generative AI [Manuscript submitted for publication]. School of Education, Capital University

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133–156. https://doi.org/10.1080/13639080020028747

Gregorcic, B., & Pendrill, A.-M. (2023). ChatGPT and the frustrated Socrates. *Physics Education, 58*(3), 035021. https://doi.org/10.1088/1361-6552/acc299

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research, 19*(1), 010132. https://doi.org/10.1103/PhysRevPhysEducRes.19.010132

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the Assessment of Non-experts' AI literacy." *Computers in Human Behavior Reports, 12*, 100338. https://doi.org/10.1016/j.chbr.2023.100338

Malik, R., Abdi, D., Wang, R., & Demszky, D. (2025). Scaffolding middle-school mathematics curricula with large language models (EdWorkingPaper No. 24-1028). Annenberg Institute at Brown University. https://doi.org/10.26300/b47y-mh41

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

Mollick, E. (2024). *Co-intelligence: Living and working with AI.* Portfolio/Penguin.

Mollick, E. R., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. arXiv preprint arXiv:2306.10052. https://doi.org/10.48550/arXiv.2306.10052

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education, 23*(1), 1–23.

Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. *School Science and Mathematics.* https://doi.org/10.1111/ssm.18356

Hoyun Cho is a full professor in mathematics education at Capital University's School of Education, where his research focuses on the role of teachers and curriculum in mathematics instruction. His work explores mathematical task design, teacher noticing and reflection, and the integration of AI in mathematics classrooms. Dr. Cho has contributed to national curriculum development as an author and researcher for South Korea's Ministry of Education.