Melese, F. M. (2026). Lesson planning with Generative AI: Developing pre-service teachers' intelligent TPACK. *International Journal of Technology in Teaching and Learning, 22(1),* 69-91.

Lesson planning with Generative AI: Developing pre-service teachers' intelligent TPACK

Fasika Minda Melese
University of Pennsylvania

Despite the growing prevalence of Generative Artificial Intelligence (GenAI) tools in higher education, educators remain underprepared to navigate the complex demands of integrating these technologies into their practice. Educators need technological, content, and professional knowledge to access and critically evaluate AI-generated content. The inability to synthesize technology with pedagogy and content knowledge risks overreliance on GenAI tools or avoidance of technology altogether. This study aims to address this gap by exploring how preservice teachers develop intelligent technological, pedagogical, and content knowledge (i-TPACK) through the integration of structured CustomGPT tools in lesson plan design activities. Using a mixed-methods research design involving 37 preservice teachers (PSTs) who are enrolled in a technology integration course at a private university in the northeastern United States, the study addressed these research questions: 1) How does using GenAI tools in lesson planning activities impact the development of preservice teachers' intelligent technological, pedagogical, and content knowledge (i-TPACK)? 2) How do preservice Teachers critically evaluate the reliability, bias, and ethical implications of GenAI-generated content for lesson planning? Data sources included pre- and post-surveys, CustomGPT prompts, and written reflections. Results show that structured GenAI integration can positively impact PSTs perceived technological, content, and pedagogical knowledge. PSTs find CustomGPT a valuable tool for brainstorming, generating ideas, and refining lesson plans, but they also raise concerns about inaccuracy, age-appropriateness, and real-world applicability. Verification strategies often involve cross-referencing, personal knowledge, prompt refinement, and peer consultation. The study contributes to the growing body of scholarship in GenAI-specific knowledge and frameworks in teacher education.

Keywords: Preservice teachers, Large language models, AI, TPACK

Dr. Fasika Minda Melese is a postdoctoral researcher at the Graduate School of Education, University of Pennsylvania, and can be reached at fasikaye@upenn.edu
DOI: https://doi.org/10.37120/ijttl.2026.22.1.05

INTRODUCTION

ChatGPT, a Generative Artificial Intelligence (GenAI) technology, has gained popularity since its release in November 2022, with over 188 million daily users as of September 2025 (Chatterji et al., 2025). In an attempt to adapt it to education, Kim and Adlof (2024) conceptualize ChatGPT as a constructivist learning tool encompassing four learning elements: providing meaningful context, supporting knowledge construction, facilitating collaborative activities, and enabling conversation with others (Kim & Adlof, 2024; Van den Berg & Du Plessis, 2023). However, as an emerging technology, ChatGPT comes with several challenges, including data privacy and transparency issues, algorithmic biases, and the potential for unethical use (McQuillan, 2022; Melese, 2025). Hallucinations (Buolamwini, 2023), for instance, pose a risk to student learning by introducing misinformation, leading to gaps in foundational content knowledge in disciplines such as science or history. In addition, critical scholarship on AI in education raises concerns about these tools and their implications for equity, inclusion, and disability (Foley & Melese, 2025; Miao & Holmes, 2023). These and other issues underscore the need to explore how educators, particularly preservice teachers (PSTs), are trained to develop the competencies required to navigate the affordances and constraints of Generative AI tools in diverse classroom contexts.

Classroom technology integration research suggests that preservice teachers integrate technology effectively when they understand how technology relates to content, pedagogy, and context, and apply these connections in meaningful ways (Lei, 2009). Thus, PSTs' integration of GenAI tools in instruction requires a deeper examination of how these technologies intersect with subject-specific content areas, effective pedagogical strategies, and the contexts in which they are applied (Petko et al., 2025). It also requires PSTs, as future educators, to ensure these technologies are used in ways that uphold the ethical standards, protect student information, and promote inclusive learning environments. Otherwise, the inability to synthesize emerging technology with pedagogy and content knowledge risks overreliance on tools (Kasneci et al., 2023) or the avoidance of technology altogether (Ertmer et al., 2012).

Technology, Pedagogy, and Content Knowledge (TPACK) is a framework designed to measure teachers' knowledge of technology integration in their classrooms (Koehler et al., 2013). Expanding Shulman's (1986) PCK model, Mishra and Koehler (2006) proposed to include the third knowledge domain, technological knowledge, for teachers to properly teach in the 21st century. However, unlike traditional digital technologies, GenAI introduces unique generative, conversational, and context-specific affordances (Mishra et al., 2023). This requires PSTs to evaluate bias, hallucinations, reliability, and ethical implications while also aligning GenAI outputs to instructional goals. Celik (2023) developed the Intelligent-TPACK (i-TPACK) to expand the traditional TPACK, considering ethics as an additional component. At the time of conducting this study, this intelligent-TPACK was the most empirically developed framework, notable for its explicit inclusion of ethics as a fifth knowledge dimension alongside the other core TPACK domains as defined below. Hence, we chose it for this study.

- Intelligent Technological Knowledge (iTK) tackles the knowledge to interact with AI-based tools and to use fundamental functionalities of AI-based tools. This component aims to measure teachers' familiarity level with the technical capacities of AI-based tools.
- Intelligent Technological Pedagogical Knowledge (iTPK) addresses the knowledge for pedagogical affordances of AI-based tools, such as personal and timely feedback, and monitoring students' learning.

- Intelligent Technological Content Knowledge (iTCK) focuses on the knowledge of field-specific AI-based tools. It aims to measure the extent to which teachers use AI-based tools to update their content knowledge. This component also focuses on teachers' understanding of AI technologies, which are more suitable for addressing subject-matter learning in their field.
- Intelligent Technological Pedagogical Content Knowledge (iTPACK) aims to measure teachers' professional knowledge to select and use suitable AI-based tools for performing a teaching strategy (e.g., monitoring and timely feedback) to accomplish the instructional goals in a specific subject-matter domain.
- Ethics aims to measure teachers' assessment of the decision on GenAI-based tools. In the context of AI, ethical competence requires educators to actively interpret, question, and assess AI-generated outputs in relation to fairness, transparency, data privacy, inclusiveness, and accountability. This emphasizes not only knowing how to use the tools, but also understanding how they work, what their limitations are, and what consequences their use may have in classroom instruction (Miao & Holmes, 2023).

The emergence of i-TPACK as a framework that considers ethics is promising for preparing educators to critically leverage GenAI tools. Despite its theoretical advances, iTPACK has a limitation that the present study aims to address. At the time of this writing, several studies (Celik, 2023; Choudhury et al., 2024; Dogan et al., 2025; Ning et al., 2024) have proposed and validated variations of AI-related TPACK scales using the generic category of "AI-based tools". However, "AI" is a broad term that includes machine learning systems, intelligent tutoring systems, adaptive learning platforms, conversational agents, etc. (Foley & Melese, 2025). This study's modification of the iTPACK scale (replacing AI terms with GenAI) represents a theoretical adaptation to address this issue.

The purpose of this study is to examine how preservice teachers (PSTs) develop intelligent technological, pedagogical, and content knowledge through structured engagement with a custom-designed ChatGPT assistant (CustomGPT). Building on emerging literature on AI literacy, prompt engineering, and GenAI-supported instructional design, this study examines how PSTs conceptualize and apply GenAI in lesson planning, how they refine AI-generated outputs using pedagogical criteria, and how their perceptions of reliability and ethics evolve throughout the intervention.

This work is grounded in the Technological Pedagogical and Content Knowledge (TPACK) framework (Mishra & Koehler, 2006) and its emerging extensions for AI-enhanced teaching, intelligent TPACK. Additionally, Maloy et al.'s (2021) critical media literacy framework informed PSTs' analysis of GenAI outputs for accuracy, appropriateness, and ethical considerations.

METHODS

This study employed a convergent parallel mixed-methods design (Creswell & Plano Clark, 2017) to examine how preservice teachers (PSTs) develop intelligent Technological Pedagogical and Content Knowledge (iTPACK) through structured engagement with a Generative AI (GenAI) tool during lesson planning. In this design, quantitative and qualitative data were collected in parallel, analyzed separately, and then merged to provide complementary perspectives addressing the following research questions:

1. RQ1: How does using GenAI tools (e.g., CustomGPT) in lesson planning activities impact the development of preservice teachers' intelligent technological, pedagogical, and content knowledge (iTPACK)?
2. RQ2: How do PSTs' interactions with GenAI influence their perceptions of reliability, limitations, and ethical considerations of GenAI content?

RQ1 was addressed through pre- and post-intervention iTPACK surveys and qualitative content analysis of PSTs' interaction prompts with the CustomGPT tool. RQ2 was addressed through post-intervention written reflections.

PARTICIPANTS AND SETTING

The study was conducted during the Spring 2025 semester of a teacher preparation course on technology integration in instruction at a large private university in the Northeastern United States. Prior to enrolling in the course, participants completed a structured sequence of coursework and field placement requirements. Table 1 shows the number of participants by gender and major. Participants include 32 females and 5 males, for a total of 37 preservice teachers from health and physical education, inclusive adolescent education, inclusive childhood education, and the inclusive elementary and special ed teacher prep program.

Participants were recruited using convenience sampling because of the nature of the course. The course met six times during the semester, covering one module each week. The course content progressed deliberately: from foundational technology integration skills to structured engagement with GenAI tools to critical and ethical evaluations of AI-generated content guided by UNESCO's guidance on generative AI in education and research (Miao & Holmes, 2023). PSTs used a CustomGPT designed with TPACK-aligned prompting structures and contextual questions (subject, grade, standards, learner needs, and available technologies) in the lesson plan revision activities.

Table 1. *Number of participants by major and gender*

Major	Female	Male	Grand Total
Health and physical education		2	2
History/Inclusive adolescent Education		1	1
Inclusive Childhood education	18	1	19
Inclusive elementary and special ed teacher prep program	14	1	15
Grand Total	32	5	37

Note. This study was reviewed and approved by the Institutional Review Board of Syracuse University (Protocol #24-282).

THE DESIGN OF CUSTOMGPT (TPACK INTEGRATED LESSON PLANNER)

Prompting in generative AI (GenAI) tools refers to the input and instructions given to Large Language Models (LLMs) to direct their output to specific tasks (Sahoo et al., 2024). The extensive datasets used to train pre-trained language models, such as ChatGPT, allow them to generate substantial linguistic and visual content. Effective prompts can therefore adapt the generated content to specific tasks such as lesson planning, assessment, etc. For this purpose, a CustomGPT was designed with structured prompts and instructions to be used to help PSTs develop TPACK-integrated lesson plans. This instructional support tool helps PSTs create, revise, and align their lesson plans with TPACK (Koehler et al., 2013), giving them an educational framework for interacting with ChatGPT when designing a lesson. The objectives of this tool are:

1. To help pre-service teachers write effective prompts when utilizing GenAI tools
2. To customize general-purpose tools, i.e., ChatGPT, to help with specific tasks (lesson planning)

The researcher designed the CustomGPT using the ChatGPT Plus 2025 version (with a paid subscription). To create it, the researcher logged in to the ChatGPT web platform, clicked Explore GPTs in the left sidebar, and clicked Create in the top right corner to open the builder interface. Then, the researcher configured the tool manually, naming it "TPACK

integrated lesson planner" and adding instructions on what it should do, what it should not do, and conversation starters as shown in the sections below.

WHAT IT SHOULD DO

Its role is to guide, suggest, and refine ideas rather than to dictate or replace the teacher's own expertise. It acts as a brainstorming partner offering structured prompts, pedagogical strategies, and technology tool recommendations that balance pedagogical goals with appropriate tools. The GPT encourages lesson planning by asking guiding questions, presenting best practices, and offering constructive feedback. Before giving any lesson plan ideas, it should also ask the user which specific state standard they are planning to use. It adapts its suggestions to the user's context. Before giving any lesson plan ideas, it should ask the user specific questions about the context, such as grade level, subject matter, learner characteristics, and existing technological tools available in the classroom. The GPT should also encourage reflection by asking questions about the generated content and give iterative support for follow-up questions. It should be collaborative with the user to improve lesson plans through multiple iterations.

WHAT IT SHOULD NOT DO

The GPT should not suggest tools, activities, or content that violate state standards, student privacy, or ethical guidelines. It should facilitate guidance instead of overriding user autonomy. It should not default to the same recommendations for every scenario; instead, it should take into account specific contextual information (grade level, subject matter, learner characteristics, and existing technological tools available in the classroom) provided by the user. It should avoid giving overly technical or complex instructions that discourage users who are less familiar with some technologies and frameworks.

PROVIDED CONVERSATION STARTERS

- Help me design a 45-minute lesson plan on Photosynthesis aligned with TPACK
- Suggest technology tools to teach fractions to elementary students
- Provide a sample assessment plan using formative and summative methods and technology tools
- Generate an activity using virtual reality (VR) for teaching 5th grade about matter and phase change

The design guidelines in Table 2 were used to design the CustomGPT and analytically mapped to TPACK subdomains (TK, PK, CK, TPK, TCK, PCK, and TPACK). In this way, the CustomGPT tool was configured to model and reinforce the prompting progression. Before generating any lesson plan content, the tool was instructed to ask PSTs a structured sequence of contextualizing questions: the state standard or curricular objectives, the grade level and subject matter, the learner characteristics of the target class, and the technology tools already available in the classroom environment. This context-gathering sequence reflects the TPACK-based lesson planning decisions (Harris & Hofer, 2009).

THE SURVEY

The survey instrument consisted of three sections. The first section collected demographic data, including whether participants have practical teaching experience. The second section gathered prior experiences with GenAI. The third section consisted of 24 items from the Intelligent TPACK items. This section of the instrument was adjusted to explicitly reference GenAI instead of general AI terminology, since the study intervention is a GenAI tool. The Intelligent-TPACK (iTPACK) instrument has 5 dimensions: Intelligent Technological Knowledge (iTK), Intelligent Technological Pedagogical Knowledge

(iTPK), Intelligent Technological Content Knowledge (iTCK), Intelligent TPACK, and Ethics. The scale includes statistically validated items corresponding to these dimensions on a 7-point Likert scale ("1: strongly disagree" to "7: strongly agree"). This framework was selected (at the time of proposing the research) as it was the only model that incorporates an ethical component into the AI-integrated knowledge.

Table 2. *Mapping TPACK domains to the CustomGPT internal prompting guidelines*

TPACK subdomain	Aligned elements from the CustomGPT guideline
Technological knowledge (TK)	Guide, suggest, and refine ideas rather than dictate Offering structured prompts Gives iterative support for follow-up questions Avoid giving overly technical or complex instructions that discourage users who are less familiar with some technologies and framework
Content knowledge (CK)	Ask the user which specific state standards they are planning to use Ask about subject matter Do not suggest content that violates state standards
Pedagogical knowledge (PK)	Offering pedagogical strategies Presenting best practices Encourages reflection by asking questions Collaborative brainstorming partner
Technological pedagogical knowledge (TPK)	Structured prompts Constructive feedback Iterative lesson improvement Avoid overriding user autonomy
Technological content knowledge (TCK)	Technology tool recommendations that balance pedagogical goals with appropriate tools Ask about existing technological tools available in the classroom
Pedagogical content knowledge (PCK)	Adapts its suggestions to the user's context Ask about grade level and learner characteristics
TPACK (integrated knowledge)	Balance pedagogical goals with appropriate tools Collaborative improvement through multiple iterations Context-sensitive recommendations

The five dimensions of iTPACK are defined as follows:

1. iTK aims to measure teachers' familiarity level with the technical capacities of GenAI-based tools.
2. iTPK aims to measure the knowledge of the pedagogical affordances of GenAI-based tools.
3. iTCK aims to measure the extent to which teachers use GenAI-based tools to update their content knowledge.
4. iTPACK aims to measure teachers' professional knowledge in selecting and using suitable GenAI-based tools to implement a teaching strategy and achieve the instructional goal of a particular domain.
5. Ethics aims to measure the teacher's assessment of the decision on GenAI-based tools.

In addition, we removed the following 3 items from the original instrument because they did not align with the research objectives.

- I can understand alerting (or notification) from AI tools to scaffold students' learning.
- I have the knowledge to select AI tools to sustain students' motivation.

- I can select various AI tools to monitor students' learning in my teaching process.

The data were collected from two sections of the course in Spring 2025. The iTPACK survey was administered via Qualtrics during the first module (pre-survey) and the 6th module (post-survey).

Participants developed an initial technology-integrated lesson plan in the first week of the course. After explicit training on prompt engineering strategies, discussing Generative AI topics, and an initial try of the CustomGPT tool, participants refined their technology-integrated lesson plan iteratively with the CustomGPT support. Their interactions with the CustomGPT were collected using the share link feature of the tool. With the refined lesson plan, participants included a written guided reflection/critique of the output using Maloy et al.'s (2021) analysis framework. The following prompts guided their written reflections:

1. Describe how you utilized prompts to create/revise your lesson plan using the CustomGPT.
2. Describe an example situation where you trusted or doubted the accuracy of GenAI output in your lesson planning. What influenced your decision?
3. What strategies do you use to verify the accuracy of GenAI outputs?
4. What challenges did you face in aligning CustomGPT-generated output with your pedagogy?
5. What ethical concerns, if any, arose in using CustomGPT for lesson planning? How did you address them?

DATA ANALYSIS

For this study, we collected and analyzed participants' pre- and post-iTPACK surveys, CustomGPT interaction prompts, and written reflections. The survey data explored the development of PSTs' iTPACK throughout the course duration. The written reflections explored their perceptions of the use of AI-generated content in their lesson design practices. Quantitative data were analyzed using R for descriptive statistics to show the difference in means of each component of the iTPACK framework and paired-samples t-tests or Wilcoxon signed-rank tests, depending on normality checks. Qualitative data, including written reflections and CustomGPT interaction transcripts, were analyzed using an iterative coding process in MaxQDA 2024. Following this preliminary analysis, a researcher-developed codebook was used, based on iTPACK and Maloy et al.'s (2021) framework. Final thematic analysis was employed iteratively to identify and code relevant themes from the written reflections.

RESULTS

To investigate RQ1 and determine how participants perceived their iTPACK before and after the intervention, pre- and post-iTPACK surveys were administered. The online survey included adapted items targeting key components of the iTPACK framework: iTK, iTPK, iTCK, iTPACK, and the Ethics component. In total, 24 Likert-scale items were used, rated on a 7-point scale. Additionally, participants responded to five multiple-choice questions addressing their practicum experiences and previous generative AI uses. These responses provided insights into participants' familiarity with GenAI tools before and after the course.

RQ1: How does using GenAI tools (e.g., CustomGPT) in lesson planning activities impact the development of preservice teachers' intelligent technological, pedagogical, and content knowledge (iTPACK)?

As shown in Table 3, many participants (n=28) first engaged with generative AI tools during their undergraduate or college education. A smaller subset of the sample reported

earlier adoption, with eight participants beginning use in high school and one participant initiating use prior to high school. Despite the variety of options for registered accounts across GenAI platforms, responses clustered in only three categories. The majority of participants (n=31) maintained accounts on ChatGPT. A small subset (n=5) identified "Other" platforms, specifically citing education-focused tools such as KhanAI, MagicSchool AI, Gemini, PopAI, Grammarly, Canva, and QuillBot as tools they have accounts on. Conversely, only one participant reported having no accounts on any GenAI platforms. These results indicate the majority of the participants have some form of experience interacting with a GenAI tool prior to taking this course and that ChatGPT serves as the primary tool for this cohort.

Table 3. *Participants' Prior Experience with Generative AI Tools (N = 37)*

	N	Percentage
First start using GenAI tools		
Before high school	1	3%
High school	8	22%
College (undergraduate)	28	76%
Never used before	0	0%
GenAI tools you have an account on (as of Spring 2025)		
ChatGPT	31	84%
I do not use any GenAI tools	1	3%
Other	5	14%
Frequency of GenAI use (pre-intervention)		
Multiple times per day	8	10.8%
Once a day	4	5.4%
A few times per week	11	14.9%
Once a week	7	9.5%
Rarely (once a month or less)	7	9.5%
Frequency of GenAI use (post-intervention)		
Multiple times per day	9	12.2%
Once a day	5	6.8%
A few times per week	12	16.2%
Once a week	6	8.1%
Rarely (once a month or less)	5	6.8%

Each scale was computed as the arithmetic mean of its respective items. As shown in Table 4, the differences in mean values for all iTPACK subscales show meaningful changes from pre- to post-survey. Before participants began the course and used CustomGPT for lesson-planning activities, their initial iTPACK survey results indicated relatively high mean scores, exceeding 4.0 on a 7-point Likert scale. Notably, the iTPACK construct recorded a mean of 4.75 with a standard deviation of 1.189. The mean intelligent technological knowledge score was exceptionally high, at 5.42, with a standard deviation of 0.835.

Table 4. *Summary of descriptive statistics for iTPACK subscales before and after intervention*

Construct	Time	Mean	SD	Median
Intelligent technological knowledge (iTK)	Pre	5.42	0.83	5.4
	Post	6.12	0.77	6
Intelligent technological pedagogical knowledge (iTPK)	Pre	5.14	1.05	5
	Post	6.16	0.80	6
Intelligent technological content knowledge (iTCK)	Pre	5.16	1.02	5
	Post	6.07	0.85	6
Intelligent technological pedagogical content knowledge (iTPACK)	Pre	4.75	1.18	4.67
	Post	6.05	0.84	6
Ethics	Pre	4.69	1.35	4.75
	Post	5.86	1.03	6

Note. N = 37.

Following, a paired-samples t-test was conducted to compare pre-intervention and post-intervention scores. There was a statistically significant increase in all the iTPACK domains. iTK changes show from pre (M = 5.42, SD = 0.83) to post (M = 6.12, SD = 0.77), $t(36)$ = -4.45, p< .001, d = 0.86, indicating a large effect. The iTPK scores showed a statistically significant increase from pre (M = 5.14, SD = 1.04) to post (M = 6.16, SD =0.80), $t(36)$ = -6.54, p< .001, d = 1.08, indicating a large effect. The iTPACK scores showed a statistically significant increase from pre (M = 4.75, SD =1.18) to post (M = 6.05, SD = 0.84), $t(36)$ = -6.88, p< .001, d = 1.24, indicating a large effect. The normality check for iTCK and Ethics deviated from a normal distribution. Therefore, we used the non-parametric Wilcoxon signed-rank test. The test showed a statistically significant increase in iTCK scores from pre (median = 5) to post (median = 6), V = 24.5, p<.001, r = 0.66. Similarly, Ethics scores increased significantly from pre (median = 4.75) to post (median=6), V = 31, p<0.001, r = 0.83.

Table 5. *Changes in Preservice Teachers' Self-Reported Intelligent-TPACK*

Subscale	Mean_Pre	Mean_Post	Mean_Diff	Test statistic	df	Sig. (2-tailed)	Effect Size
iTK	5.42	6.12	0.7	t = -4.45	36	0.0000806	d = 0.864
iTPK	5.14	6.16	1.02	t =-6.54	36	0.000000134	d = 1.08
iTCK	5.16	6.07	0.91	*V = 24.5*	36	<.001	*r = 0.66*
iTPACK	4.75	6.05	1.3	t =-6.88	-	0.000000047	d = 1.24
Ethics	4.7	5.86	1.16	*V =31*	-	<.001	*r = 0.83*

Note. All p<.001.

Quantitative analysis findings show that structured GenAI integration can positively impact PSTs' perceived intelligent technological and pedagogical skills. The largest gains occurred in the intelligent TPACK (iTPACK) subscale, likely due to the hands-on exposure to CustomGPT tool for lesson planning and other GenAI tools explored throughout the course period. Notably, AI ethics-related competencies also improved, suggesting that reflective components of the course and class discussions about the impacts of integrating AI in lesson planning supported critical engagement.

PATTERNS IN PROMPTING CUSTOMGPT FOR LESSON PLANNING ACTIVITY

During the lesson revision activity, 37 links of interaction with CustomGPT were collected. Eight participants did not provide a link, or the link provided was not functional. Five participants provided links that only showed the output and not the prompts used. These were treated as having 1 prompt. This exclusion represents a meaningful limitation for the prompt analysis. Those PSTs without a link or a non-functional link may have engaged

differently with the CustomGPT; therefore, the reported patterns in this section should be interpreted as descriptive of engagement among those PSTs who provided complete prompts. Interaction patterns were examined at the structural level by identifying how many prompts were used in each interaction. Table 6 shows the distribution of prompt frequency across participants. As shown, the majority of the participants (n=16) engaged with the CustomGPT through only one or two prompts during the lesson plan revisions. A smaller group (n=7) engaged in extended exchanges of five or more prompts. This pattern shows that PSTs used CustomGPT as a targeted revision tool rather than as an iterative co-design partner. This pattern may reflect the course design, as PSTs already have the first lesson plan created and are only asking for revisions based on that plan. This pattern may also reflect the early stages of prompt engineering development among PSTs. The capacity for iterative, reflective prompting cannot be assumed and suggests more structured practice.

Table 6. *Distribution of Prompt Frequency*

Number of prompts utilized	Total number of students (section 1)	Total number of students (section 2)	Total number of students
1 - 2	12	4	16
3 - 4	3	3	6
5+	2	5	7
N/A (link not provided)	2	6	8
Total	19	18	37

RQ2: How do PSTs' interactions with GenAI influence their perceptions of reliability, limitations, and ethical considerations of GenAI content?

To better understand how preservice teachers evaluate GenAI-generated content for lesson planning and the strategies they use to verify, participants' written reflections were analyzed using MAXQDA 24. An initial deductive coding framework was developed prior to analysis, grounded in the iTPACK and the analysis framework (Maloy et al., 2021). However, consistent with thematic analysis principles (Braun & Clarke, 2012), the coding framework remained open to the discovery of new codes not anticipated by deductive categories. The section below provides descriptions and examples of major themes that developed from this analysis. Participant names are replaced with pseudonyms.

THEME 1: PSTs WERE ABLE TO IDENTIFY CHALLENGES REGARDING GENAI USE IN LESSON PLANNING, UTILIZING THEIR TPACK AND PRIOR EXPERIENCES IN THE CLASSROOM

PSTs express concerns regarding the reliability and accuracy of GenAI outputs, particularly when the suggestions diverge from their professional judgment or known educational contexts (e.g., misaligned educational standards). Doubts arise when the CustomGPT tool outputs activities that are too lengthy or complex for the allotted lesson length or target age group, lack student-centered or hands-on activities, insufficient differentiation or inclusivity, and unrealistic activity loads. For example, Andy doubted the accuracy of Generative AI output when it suggested adding an exit ticket or Google Form to a lesson plan, which would have taken too much time. This decision was influenced by the fact that Andy had not specifically asked for adjustments to the lesson's closing, and the AI suggested lengthy activities.

> "One time I doubted the accuracy of Generative AI output was some of the things it asked me to add into my lesson. I believe it gave me options for an exit ticket or google form which would have taken up too much time in my initial lesson plan. I decided to question the output as I had not specifically

> asked for adjustments to be made on my closing of the lesson, and it suggested lengthy activities." (Andy)

To verify this, Andy used his personal knowledge of the standard and available time frame to evaluate if the information aligns with the standard and the ability to implement the lesson. This strategy proved effective, resulting in a lesson plan that aligned with the standard and was reasonable within the time frame. Similarly, Leslie relies on her understanding of the content and contextual understanding when GenAI output doesn't align with her pedagogical approach. Her decision to question the output was influenced by the fact that the NYS standard's definition of persuasive writing differed from that of argumentative writing, and GenAI had swapped the two definitions. She used information observed in her placement and education, demonstrating her contextual understanding in evaluating the usefulness of the output she got from GenAI.

> "the few times I used GenAI to help me lesson plan I questioned the accuracy of the information. Over the past few weeks, I have been creating a lesson on persuasive writing. I wanted to make a lesson where students broke down the information that needed to be incorporated into a compelling essay's introduction paragraph. I asked GenAI about the type of writing and started changing the name to an argumentative piece. This confused me, as it changed the topic of the activity. One fact that led me to trust or question the output was that the definition in the NYS standard of persuasive writing differed from argumentative writing. GenAI had swapped the two definitions. I also used information that I observed in my placement and education. As a result, I stopped using GenAI as a tool to provide me with information about definitions." (Leslie)

Similarly, drawing on her prior classroom experience, Myra exercised professional judgment by questioning the accuracy of GenAI's specific timing estimates for activities. For instance, the tool suggested 6 minutes for first graders to segment and blend 5 words in groups, a metric Myra did not trust based on her experience with her students' capabilities, transition times, and the demands of group collaboration. Myra emphasized that personal experience with children in a classroom setting makes a teacher a better indicator of task duration.

> "At one point in the lesson, the AI tool said that I would need 6 minutes to have children do segmenting and blending of 5 different words in their groups. I did not trust this metric by ChatGPT, as doing 5 words in 6 minutes is already a lot to ask of 6- and 7-year-olds, especially when they are in a group setting and asked to collaborate. Additionally, 1st graders do still take time to transition and to log into their devices which would be necessary to complete the second part of my lesson. To me, it felt obvious that when it came to specific timing, I was a better indicator of how long certain tasks will take children, because I have experience being with these children in a classroom setting, and understanding what they are capable of in a specific time frame." (Myra)

Collectively, these reflections suggest that preservice teachers, when engaged in experimentation with GenAI and reflective practices, use their TPACK and prior experience to navigate the challenges in aligning with their pedagogical approaches. In other words, pedagogical tension with GenAI output enforces the enactment of agency, leading to reflective lesson planning.

On the other hand, some participants trust AI outputs when they align with familiar pedagogical practices, previously used resources, or established standards. For example, for Adele, CustomGPT significantly impacted the quality and efficiency of the lesson plan. She states that the tool provided a full evaluation of the lesson plan, offering constructive

feedback, strengths, and areas for improvement, which allowed for the addition of more in-depth information and differentiation to benefit students. CustomGPT suggested incorporating a guided think-aloud to increase student understanding of weather prediction, which Adele initially doubted but later recognized as a valuable and quick activity to enhance student comfort with technology.

> "An example of a time I trusted the accuracy of the Generative AI was when it suggested me to increase students understanding through a guided think-aloud, in which the teacher would model the prediction of weather based on the colors or symbols on the weather map, to show the students what their thought process should look like. At first, I was not sure that this was necessary in my lesson plan but upon further reflection I realized that this is a great tool and activity to do quickly with the class to get them more comfortable with the technology." (Adele)

Similarly, suggestions for age-appropriate activities and technology tools, especially when aligned with known educational resources, fostered trust. For example, Claire requested stronger technology resources in her prompt ("make the technology tools stronger"), leading to the suggestion of specific websites and videos that enhanced the lesson plan's materials. This process improved the quality of the lesson plan's content. Claire trusted the accuracy of the GenAI's suggestions, particularly when the tool recommended Brain Pop Jr., a known and trusted resource.

> "I trusted the accuracy of when I asked for stronger technology materials to use. I trusted it because, being a science lesson, it gave me Brain Pop Jr. as a tool to use, a website that I know is trusted and that I have used in the past. This is what led me to trust this output. However, I didn't blindly trust it, I ensured that the material still existed and fit the objectives I was looking to hit with that material. Overall, the CustomGPT gave me strong tech examples." (Claire)

Another example is Eva, who found that CustomGPT helped her think more critically about integrating technology into her lesson plan, supporting teaching goals, and engaging students. The tool guided her in combining technology such as Google Slides, the Bugs and Numbers app, which she knows is age-appropriate and aligned with the lesson objectives.

> "I trusted more of the suggestions from CustomGPT, especially when it came to using the Bugs and Numbers app and Google Slides. These tools seemed perfect for my goal of teaching students how to count to 20 and understand one-to-one correspondence. The app is age-appropriate and focuses directly on counting, which is aligned with my lesson objectives." (Eva)

Jewel trusted CustomGPT's suggestion, recognizing its past effectiveness in promoting inclusivity in classrooms. For instance, it suggested adding text-to-speech and translation extensions, which were trusted by her due to prior positive experience and alignment with promoting inclusivity and accessibility in the classroom.

> "During my lesson revision with CustomGPT, I asked it to make the lesson inclusive and effective. CustomGPT then suggested I add in a text-to-speech and translation extension. I trusted the AI's suggestion as it was something that I have used so many times and have seen benefit students in many ways." (Jewel)

Overall, PSTs navigated challenges in aligning GenAI output with their lesson plan based on their developing TPACK and prior classroom experience.

THEME 2: PSTs MOVE BEYOND FUNCTIONAL TOOL USE AS THEY ENGAGE IN A PROCESS OF ETHICAL EVALUATION AND STRATEGIC RESISTANCE

PSTs raised several ethical concerns while utilizing CustomGPT for lesson planning, such as bias in content, misinformation, generic suggestions, a lack of contextual understanding, data privacy, and over-reliance on GenAI. Despite these challenges, the majority of PSTs chose to use GenAI as a supplementary tool, selectively adopting, anonymizing, and limiting data, reflecting on their teaching, monitoring student inputs, and preserving the human teacher.

Bias in content. This includes AI reflecting biases from its training data, leading to culturally insensitive, Eurocentric, or stereotypical content. For example, Jan mentioned AI's tendency to depict doctors as white men:

> "I think that an ethical dilemma with using Chat GPT is that it is not programmed to be cognizant of social contexts. It can only understand the resources that have been plugged into it. It is vulnerable to reflecting historical biases. For example, in one of my past classes when asked to draw a picture of a doctor, the AI would only produce white men because that is the stereotype for who the most doctors are which is not true." (Jan)

Misinformation. Some participants have encountered factual inaccuracies, such as GenAI tools generating non-existent articles (as shown in the quote below by Julian) or misrepresenting definitions, which led them to question the tool's ability to provide contextually appropriate or up-to-date information. The potential for ChatGPT to provide outdated information, such as recommending Jamboard despite its discontinuation, also fuels skepticism.

> "One time, I asked ChatGPT to find me some articles I can use for my research paper, and it generated articles that didn't exist. So, that is the downside of using ChatGPT because it is not always reliable and up to date with modern information. ChatGPT was unable to give me articles that existed and it made me doubt the accuracy of GenAI." (Julian)

Generic suggestions and lack of contextual understanding. Several PSTs also highlighted that AI cannot understand the unique dynamics, individual needs, or personal lives of students in a classroom. For example, Geneva acknowledges that AI can lack contextual understanding, not knowing past lesson plans, student backgrounds, or individual student needs unless explicitly provided.

> "I think that with GenAI there can be a lack of contextual understanding. AI will not know all of the lesson plans that have been taught in the past (unless it is given to it), or the background of your students, or each student's needs etc. Although a lot of this information can be put in, it honestly may be less work than creating a lesson plan yourself." (Geneva)

Similarly, Leslie identified ethical concerns when using CustomGPT for lesson planning, noting that AI does not understand the diverse needs of students in a classroom, as it lacks the human element of daily interaction and awareness of students' personal lives, potentially leading to lesson plans that overlook important considerations such as students' actual cognitive level and environmental factors around the classroom.

> "GenAI does not understand the diverse needs of the students in the classroom as it is technology and does not spend every day teaching each student. GenAI may leave out important considerations when creating lesson plans, like what is going on in students' personal lives. Next, CustomGPT may make the lesson plan too advanced. If I were to give CustomGPT a prompt about a lesson plan and a standard, it may go beyond a student's skill set. Lastly, CustomGPT may struggle to make an assessment that includes

> all students in the classroom. CustomGPT may fail to acknowledge student struggles due to environmental issues, such as how students may be distracted due to noise." (Leslie)

Zera notes that AI outputs can be generic and may not account for students' diverse cultural backgrounds, which conflicts with the pedagogical approach of culturally relevant pedagogy (Ladson-Billings, 1995). This means GenAI outputs often fail to ensure all learners feel honored and represented, requiring PSTs to manually add these elements to lesson plans.

> "For example, when I was revising my first lesson plan for this course I noticed that GenAI is very generic and is not aware of each overall classroom dynamic and makeup of each classroom. For example, in my placement last semester in Block II, I had 5 ENL students, some were Chinese, Japanese, Korean, and Russian. All of these ethnicities GenAI did not take into consideration when revising my lesson plan to cater to all students to make each student feel honored or welcomed which is part of my teaching philosophy." (Zera)

PSTs also mentioned concerns regarding GenAI tools' inability to account for specific classroom resources and student demographics, which are essential for successful lesson planning. However, they also expressed concerns that explicitly providing background information to the CustomGPT creates privacy concerns associated with sharing sensitive information with the GenAI tool. This creates a tension between what to share and what not to share. It also adds an additional AI safety and security layer for PSTs to have as a skill. For instance, Zera notes that AI does not recognize crucial classroom specifics such as physical appearance, available resources, or the student demographics, which are vital for creating successful lesson plans. For instance, CustomGPT might assume all students have access to individual school-owned Chromebooks, which is not always the case due to have in different schools.

> "For example, in each of my lesson plans that I have created for my field placements I have had to vividly describe the background of the school I was placed in, the background of my classroom, and the makeup of the students in my classroom. This helps paint a clear picture in the audience's mind of what resources I have available as each school district and classroom are widely unique. Thus, CustomGPT does not take these factors into consideration when lesson planning which is difficult as some lessons are created assuming that each student has access to their own school-owned Chromebook which is not always the case due to financial reasons in each school. Thus, in order to include these ethical considerations into my lesson plan I made sure to read what CustomeGPT computed to me and then either tell CustomGPT the background of my classroom or physically type in the information myself. My one fear with telling AI any background information is that it is an invasion of privacy and that can be very unsafe." (Zera)

Data privacy. A significant ethical concern is the sharing of personal or identifiable student information with the GenAI tool, which could constitute a security breach. For example, Ellis noted that although no personal data was provided to CustomGPT, she "made sure to avoid including student names, IEP details, or specific demographics" in prompts.

> "I also considered the privacy of student data. Even though I didn't provide any personal data to CustomGPT, I made sure to avoid including student names, IEP details, or specific demographics in my prompts. I carefully reviewed all AI-generated content to address these concerns, adjusting ensure it used" (Ellis)

Similarly, Myra identified a major ethical dilemma when using ChatGPT for lesson planning, which is the necessity of sharing a significant amount of information about the

classroom and children to achieve the desired lesson plan standard. While specificity is important when interacting with ChatGPT, it is crucial to understand that sharing all students' information with the application can be a huge security breach and unethical.

> "One major ethical dilemma I've learned with using ChatGPT to form lesson plans is that in order for the lesson to be made to the standard that you need it to, a lot of information will have to be shared about your classroom and the children in it. However, sharing a lot of personal, identifiable information about young children to an online platform proves to be a huge security breach, as well as wildly unethical." (Myra)

Over-reliance and Intellectual Property (IP) concerns. Several participants worry about becoming overly reliant on GenAI and plagiarism when using GenAI for lesson planning. For example, Andy identified this concern as the potential for theft of intellectual property if one were to use the generated lesson exactly as provided. He also expressed concern about not knowing the IP's original source.

> "While using GPT for lesson planning, the biggest ethical concern to me would be using the lesson exactly as it gave it to you. I think this is some kind of theft of someone's intellectual property, and the fact you don't know who's is even scarier." (Andy)

While acknowledging AI as a powerful tool for effectiveness and materials, PSTs emphasize that a teacher's role is to create and design effective lessons without over-relying on AI. Similarly, Simona identified that people might rely too much on AI to create things for them, potentially leading to a loss of their own abilities to perform these tasks.

> "An ethical concern may be that people rely too much on AI to create things for them, and we loose abilities to perform these tasks on our own. AI may not always be the most culturally relevant or inclusive as well which can cause many ethical dilemmas." (Simona)

Rather than abandoning the technology in response to these concerns and issues, many PSTs demonstrate teacher agency by proposing and enacting specific mitigation strategies, as further discussed below. This process is what leads them to develop Intelligent Technological Pedagogical Content Knowledge (iTPACK), where their professional judgment mediates algorithmic logic. This evolution from anxious skepticism to proactive management suggests that the development of Critical AI Literacy, as the ability to understand, use, and critically assess AI output, should be a prerequisite for technology integration in education.

STRATEGIES TO ADAPT OR RESIST GENAI OUTPUT

To address concerns about accuracy and ensure reliability, PSTs employ various verification strategies such as cross-referencing with trusted sources, applying personal and professional judgment, peer consultation, and specific prompting techniques.

Cross-referencing with credible sources. This was the primary method to verify GenAI outputs among PSTs. They mentioned checking information against academic articles, reputable educational websites (e.g., Common-Sense Education), and curriculum standards (e.g., NYS standards). For instance, when the AI tool provided detailed information about specific weather types for a first-grade lesson plan, Adele questioned its accuracy because the information was more detailed than what would be taught to first graders. The strategy was to cross-reference academic articles and to ensure that the simplified information provided to students was still correct.

> "As for assessing and verifying the accuracy of Generative AI outputs, I would usually reference an academic article that has the correct information about the topic. When looking at my lesson plan about weather, I was given information about the specific weather types from the GenerativeAI, that I

> was not sure was completely accurate. The GenerativeAI gave much more detailed information about the weather types than I would be teaching my first-grade students, so I wanted to make sure that the more simplified information I was providing to my students was still correct." (Adele)

Ellis cross-referenced the GenAI's recommendations against real-world data, such as verifying map questions and scenarios against a real map of food access in Syracuse to ensure contextual correctness.

> "Since the GenAI output was in line with the learning goals specified in the NYS Social Studies Framework 3.9a. The language used was suitable for 3rd graders, and the AI-generated scenario promoted empathy and critical thinking. Furthermore, I verified the map questions and scenario against a real map of food access in Syracuse to ensure that the AI's recommendations were contextually correct." (Ellis)

Pedagogical reasoning was also used, involving self-questioning like "Would this make sense for a third-grade learner?" This helped customize AI recommendations. Additionally, PSTs occasionally sought feedback from a friend to review GenAI-generated content when doubts arose, finding this effective for gaining different perspectives on what could be improved or what worked well.

Follow-up prompts and iterative refinement. When initial AI outputs are too general, complex, or misaligned, PSTs refine their prompts to be more explicit. Many PSTs find that by providing detailed prompts or clearly outlining their teaching values and strategies, CustomGPT generates suggestions that align well with their pedagogical approach. This is particularly true when seeking ideas for differentiation, inclusivity, or specific learning styles. For example, Jewel ensured that her prompts clearly outlined her teaching values and strategies, specifying that activities needed to be engaging, fun, inclusive, and interactive. This detailed and intentional approach from the beginning resulted in AI outputs that were accurate and easy to implement, leading to a few challenges in aligning the outputs with their pedagogical approach.

> "Through my experience using CustomGPT, I utilized prompts to refine my lesson plan to be focused on cultural awareness and inclusivity. First, I shared the overall theme of the lesson and then continued to follow up with more detailed prompts like my learning objectives, student needs, and assessments. I believe that CustomGPT responded really well to this form of interaction; it was able to provide me with support in shaping my ideas and taking them a step further while maintaining a connection to my audience." (Jewel)

Geneva also employed prompt engineering by specifying the desired lesson plan length, learning objective (which AI revised), grade level, and available student materials. She also provided a link to a game and asked the AI how to best integrate it into the lesson plan.

> "I think the most important thing is to revise your prompts as you go and to specify EXACTLY what you are looking for. This helps ensure that the output is what you want, rather than random or unneeded information. Something that I think helped me was first giving the GenAI my learning objective. Like a real person, when planning a lesson plan, it's important that the lesson is based on the objective. That way, AI had a good understanding of the goal before it started to create plans and activities to relate to the objective." (Geneva)

Overall, PSTs initially asked CustomGPT for help incorporating more technology into a lesson plan. Sometimes, the tool's response was helpful and guided planning by highlighting forgotten elements like quizzes and differentiation strategies, such as adding captions to videos. But specific issues were encountered when the initial prompts were not specific

enough. To resolve this, PSTs made sure subsequent prompts were more specific about what was being sought. And they also reviewed the AI output to better fit specific needs.

Rejection of misaligned suggestions. Participants did not hesitate to reject AI suggestions that contradict their core pedagogical beliefs, are impractical in their classrooms, or do not meet the learning objectives they set out. This shows the development of their pedagogical reasoning through interacting with GenAI tools for lesson planning (Celik et al., 2026). For example, Leslie faced challenges aligning CustomGPT-generated output with her pedagogical approach, leading her to reject many suggestions. One specific issue was when CustomGPT suggested adding math to an ELA vocabulary lesson, which went against the lesson's subject matter. Another challenge arose when CustomGPT proposed activities that would extend the lesson plan beyond its fifteen-minute time limit. To resolve these challenges, Leslie considered her students, the curriculum, and past activities, rejecting suggestions that conflicted with curriculum expectations or exceeded students' skill sets within the given time frame.

For Viola, there was a discrepancy between her plan and the AI tool's suggestion to include technology in every step of the instructional sequence. She did not want students to rely on technology for everything, particularly for exit tickets. For instance, the AI suggested using platforms like Flipgrid or Padlet for a 1-minute exit ticket, which she deemed infeasible as it presumed students were already accustomed to these platforms and could access and use them with ease. Her pedagogical goal for technology was to engage students and enhance efficiency, but CustomGPT's directions did not consider this.

For the above-mentioned reasons and others, several PSTs emphasize using GenAI as a brainstorming or idea generator rather than as a replacement for their own creativity and expertise. More views focused on the CustomGPT as a resource rather than a creation tool, emphasizing that lessons should be the PSTs' own words, using the tool for feedback and suggestions purposes. The primary goal in integrating CustomGPT was to enhance instructional flow, differentiate strategies, and better align with TPACK principles, including clarifying learning objectives and connecting them to assessment methods. This centered the human educator in the lesson planning. PSTs identified GenAI lacks empathy and a personal touch, which could lead to a loss of human connection crucial for social-emotional learning if relied upon too heavily.

Although there is a strong push to treat generative AI tools as collaborative partners, the evidence shows its impact is surface-level. This limitation becomes even more pronounced when considering the *trade-offs* in ethical and professional judgment. As a technological tool, AI creates this tension among PSTs: the line between using it and crossing a security threshold is thin. Most PSTs mentioned they would choose the ethical mitigation plan for all their ethical concerns, but without the foundational ethical knowledge, the outcomes are highly unpredictable. Therefore, developing PSTs' professional ethics and critical thinking skills in teacher education programs is crucial.

SUMMARY OF FINDINGS

SIGNIFICANT GROWTH ACROSS ALL iTPACK DIMENSIONS

To assess how participants perceived their iTPACK before and after the intervention, we administered pre- and post-iTPACK surveys. The online survey included adapted items with modifications for Generative AI terminology, targeting key components of the iTPACK framework: iTK, iTPK, iTCK, iTPACK, and Ethics. The pre- and post-intervention survey results demonstrated statistically significant increases across all five dimensions (iTK, iTPK, iTCK, iTPACK, and Ethics) of the iTPACK framework with large effect sizes. These findings suggest that structured engagement with a TPACK-aligned CustomGPT tool within a course-based intervention can meaningfully enhance PSTs' self-

perceived competencies in integrating GenAI with pedagogy, content, technology, and ethical awareness. The largest gain was observed in the iTPACK construct, followed by iTPK. These results indicate that the intervention effectively developed PSTs' self-perceived integrated knowledge of how GenAI tools can align with pedagogical strategies and content goals, a core premise of the iTPACK framework.

The growth in the Ethics dimension also suggests that hands-on engagement with GenAI, coupled with reflective activities and discussions, can foster ethical awareness. These findings align with Antonio (2025), who found statistically significant improvements across all AI-TPACK dimensions among preservice science teachers following a course-based intervention. Although in different contexts, this pattern suggests that structured, pedagogically grounded AI integration courses can reliably develop PSTs' self-reported AI-related TPACK. These findings also align with Celik's (2023) original iTPACK study, which established that teachers' technical familiarity with AI tools (iTK) underpins the development of higher-order integration knowledge (iTPK and iTPACK).

In this study, iTK had the highest baseline mean (5.42), suggesting participants entered the course with relatively strong general familiarity with GenAI tools, which is likely due to the widespread adoption of ChatGPT among college students, as 84% of participants noted having a ChatGPT account, and 76% first engaged with GenAI during their undergraduate education. Hence, the intervention built on this foundational technological knowledge to develop more situated integration. Participants in this study, predominantly sophomores and juniors who completed a practicum in a local school before/during this course, demonstrated substantial growth in iTPACK. These findings also align with other studies that found students with active training or less teaching experience showed enthusiasm and readiness for AI integration (Baena-Morales et al., 2025).

The thematic analysis of PSTs' interaction with CustomGPT provided insight into how they engaged with GenAI prompting. Most participants used only one or two prompts, and the content of these prompts was predominantly revision-oriented. No participant used CustomGPT to generate an entirely new lesson plan from scratch. This pattern of use reflects what Celik et al. (2026) described as co-constructive engagement, where the human designer retains ownership of the lesson structure while leveraging AI for targeted enhancement. However, the relatively shallow depth of prompts (most use only 1-2 prompts) raises questions about whether the tool is fully exploited iteratively. This may be due to time constraints (the 15-minute lesson format), unfamiliarity with iterative prompt engineering strategies, or the perception that minimal interaction was sufficient for the task. Future research might explore iterative prompting strategies and encourage deep engagement cycles.

CRITICAL EVALUATION OF GENAI CONTENT

To better understand how preservice teachers evaluate GenAI-generated content for lesson planning and the strategies they use to verify it, we analyzed participants' written reflections; the following shows key findings for RQ2.

Identifying challenges through TPACK and prior experience. PSTs demonstrated an ability to identify challenges in GenAI-generated content, drawing on their developing TPACK and prior field placement experiences. They identified challenges such as misalignment with time constraints, inaccurate content, and unrealistic activity expectations. Their existing content knowledge and contextual knowledge helped them identify these challenges. PSTs with stronger content knowledge detected inaccuracies more effectively, while those with classroom experience were more attuned to practical implementation issues. This pattern supports that TK alone is not sufficient for effective AI integration (Celik, 2023) and that all forms of TPACK knowledge (content, pedagogy, technology) collectively shape PSTs' ability to work with AI tools (Oved & Alt, 2025).

Moving beyond functional use to ethical evaluation. The written reflections showed that PSTs moved beyond functional engagement with GenAI to engage in substantive ethical evaluation. The ethical concerns they raised were multifaceted, including bias in content, misinformation, data privacy, over-reliance, intellectual property, and lack of contextual understanding. These observations show that PSTs did not simply accept GenAI outputs at face value but interrogated the implications of GenAI-generated content, an essential competency for educators committed to culturally responsive pedagogy (Ladson-Billings, 1995) and implementing just-in-time, justice-oriented AI adoption in classrooms (Budhai & Heath, 2025).

Strategic responses to adapt and reject. PSTs demonstrated a range of strategic responses to adapt or reject GenAI-generated content, such as cross-referencing with credible sources, iterative prompting and specificity, outright rejection, and prioritizing human connection. This progression between skepticism and strategic engagement shows that critical AI literacy develops as PSTs encounter challenges, evaluate them, and develop responses. Experiential engagement with GenAI is essential to developing the competencies future educators need.

DISCUSSION AND CONCLUSION

This study provides empirical evidence on how GenAI can support PSTs' lesson planning processes when paired with explicit instruction in prompt engineering and critical evaluation. The pre- and post-intervention iTPACK survey analysis showed that there was a major difference in survey score means from pre- to post-intervention. The survey analysis also showed statistically significant increases across all five constructs (all $p < .001$). Three of the five constructs (iTK, iTPK, and iTPACK) showed statistically significant differences in mean scores ($p<.001$) with large effect sizes (Cohen's $d = 0.86$, 1.08, 1,24, respectively). For the two constructs where difference scores deviated significantly from normality (iTCK and Ethics), the Wilcoxon signed-rank test similarly indicated large effects ($r = 0.66$ and 0.83, respectively). Thus, across both parametric and non-parametric analyses, all five domains of the iTPACK framework showed statistically significant improvements with large effect sizes following the course's intervention and discussions. This aligns with previous studies that showed that students with active training or less teaching experience (e.g., pre-service teachers) show enthusiasm and readiness for AI integration in their classrooms (Baena-Morales et al., 2025). This increase in iTPACK domains also might be due to participants' prior experience and familiarity with GenAI tools.

PSTs' experiences with CustomGPT's accuracy and pedagogical alignment significantly shape their perceptions of its reliability and their strategies for ethically integrating it into lesson planning. The written reflection findings show two dimensions of how PSTs critically evaluate GenAI output: first, PSTs' capacity to identify challenges in GenAI output, including inaccuracies, contextual misalignment, and developmental inappropriateness, drawing from their developing TPACK and prior classroom experiences; and second, PSTs' move beyond functional use to ethical evaluation and strategic resistance, in which they identified concerns about bias, misinformation, data privacy, and over-reliance. They responded through cross-referencing, iterative prompting, rejection, and the preservation of their own judgment. While many PSTs find CustomGPT valuable for brainstorming, generating ideas, and refining lesson plans, concerns about accuracy, including factual inaccuracy, age-appropriateness, and real-world applicability, lead to a cautious approach.

The findings show several practical implications for integrating GenAI into teacher education curricula. Teacher education programs seeking to integrate GenAI should invest in designing structured interaction protocols rather than simply providing access to generic AI tools. This supports research showing that facilitating conditions increase effective AI

use in pedagogy (Kgosi et al., 2026; McBean & McBean, 2025). PSTs need scaffolding to interpret and enact GenAI integration, and incidental learning through public media is insufficient to develop the critical engagement skills AI literacy demands (Velander et al., 2024). This research project's CustomGPT intervention represents one approach to providing that scaffolding within a teacher preparation context.

The ethical concerns and verification strategies that PSTs developed are rooted in direct experience with GenAI limitations, not from instruction alone. Therefore, teacher education programs should create opportunities for PSTs to encounter, evaluate, and respond to GenAI output within authentic instructional design tasks. PSTs with field practicum experience were better positioned to identify contextual misalignments in GenAI outputs. This suggests that GenAI integration activities may be most productive when sequenced after or alongside field placement experiences, allowing PSTs to draw on classroom realities as evaluation criteria.

The variation in prompting depth and quality across participants suggests that many PSTs would benefit from explicit instruction in formulating effective prompts, including strategies for specificity, iteration, and contextual framing. This aligns with the literature, which positions prompt engineering as a core competency for educators (Meskó, 2023; Sahoo et al., 2024). The process-level account of how PSTs exercise agency during GenAI-mediated lesson planning shows that PSTs approached the tool as a revision tool rather than a lesson generator. This finding extends the work on co-constructive engagement with GenAI (Celik et al., 2026) by documenting the specific design decisions and rationales that constitute agentic interaction. Future research should explore iterative prompting strategies and encourage deep engagement to better align analysis of prompt interaction with iTPACK development.

This study advances the ongoing conversations about the role of GenAI in teacher education, offering a grounded and actionable perspective on preparing future educators for AI-supported classrooms. The results align with emerging studies (Celik et al., 2026; Wang et al., 2024) and suggest that integrating GenAI into TPACK-based instruction holds promise. This study also contributes to the growing body of scholarship in AI-specific frameworks in teacher education. Existing TPACK models acknowledge the importance of foundational competencies, yet empirical research examining how PSTs actually learn to work with generative AI systems in lesson design remains limited. By analyzing real interaction logs and written reflections, this research provides evidence of how PSTs negotiate the affordances and constraints of GenAI tools in their teaching practices.

LIMITATIONS

Several limitations should be considered when using these findings. The sample size (n=37) and the use of convenience sampling limit the generalizability of the findings. All participants were enrolled in the same technology integration course at one institution, and the sample was predominantly white females with a concentration in inclusive childhood education and special education programs. The results might vary with PSTs from other disciplines, institutions, or demographic backgrounds. Another limitation is the iTPACK survey, which may introduce bias through self-reporting. This is a persistent methodological concern within the broader TPACK literature and concerns the extent to which self-report instruments capture enacted professional knowledge versus beliefs about that knowledge (Chai et al., 2013; Willermark, 2018). In addition, due to the non-experimental design and the absence of a control group, the findings should be interpreted with caution. The observed gains in the scales may be influenced by factors beyond the intervention.

REFERENCES

Antonio, R. (2025). Fostering preservice science teachers' AI-TPACK competence and reflections through an AI-focused pedagogical learning course. *Journal of Technology and Science Education, 15*(3), 784–809. https://doi.org/10.3926/jotse.3693

Baena-Morales, S., Sánchez-Jarque, M., Ferriz-Valero, A., & Bofill-Herrero, A. (2025). The paradox of experience: Generational analysis of AI integration in physical education using the iTPACK framework. *Physical Education and Sport Pedagogy*, 1–22. https://doi.org/10.1080/17408989.2025.2536145

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Budhai, S. S., & Heath, M. K. (2025). *Critical AI in K-12 classrooms: A practical guide for cultivating justice and joy.* Harvard Education Press. https://hep.gse.harvard.edu/9798895570180/critical-ai-in-k-12-classrooms/

Buolamwini, J. (2023). *Unmasking AI: My mission to protect what is human in a world of machines*. Random House. https://www.amazon.com/Unmasking-AI-Mission-Protect-Machines/dp/0593241835

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

Celik, I., Kontkanen, S., Laru, J., & Dalyanci, A. A. (2026). Co-constructing adaptive lesson plans with GenAI: Pre-service teachers' Intelligent-TPACK and prompt engineering strategies. Computers & Education., 241, 105485. https://doi.org/10.1016/j.compedu.2025.105485

Chai, C. S., Koh, J. H. L., & Tsai, C.-C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society, 16*(2), 31–51. https://www.jstor.org/stable/jeductechsoci.16.2.31

Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., & Wadman, K. (2025). *How people use ChatGPT* (NBER Working Paper No. 34255). National Bureau of Economic Research. https://doi.org/10.3386/w34255

Choudhury, S., Deb, J. P., Pradhan, P., & Mishra, A. (2024). Validation of the teachers AI-TPACK scale for the Indian educational setting. *International Journal of Experimental Research and Review, 43*, 119–133. https://doi.org/10.52756/ijerr.2024.v43spl.009

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research*. Sage publications. https://lccn.loc.gov/2017037536

Dogan, S., Nalbantoglu, U. Y., Celik, I., & Agacli Dogan, N. (2025). Artificial intelligence professional development: A systematic review of TPACK, designs, and effects for teacher learning. *Professional Development in Education, 51*(3), 519–546. https://doi.org/10.1080/19415257.2025.2454457

Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education, 59*(2), 423–435. https://doi.org/10.1016/j.compedu.2012.02.001

Foley, A., & Melese, F. (2025). Disabling AI: Power, exclusion, and disability. *British Journal of Sociology of Education*, 1–22. https://doi.org/10.1080/01425692.2025.2519482

Harris, J., & Hofer, M. (2009). Instructional planning activity types as vehicles for curriculum-based TPACK development. *Research Highlights in Technology and Teacher Education*, 99-108. https://tpack.org/lats/HarrisHoferTPACKDevelopment.pdf

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

Kgosi, A. M., Botes, W., Neethling, S., Mahlo, L., & Zondi, M. F. (2026). A TPACK-informed study on equipping pre-service teachers with digital tools, artificial intelligence and open educational resources: Integrated strategies in STEM education. *African Journal of Teacher Education and Development, 5*(1), Article 12. https://doi.org/10.4102/ajoted.v5i1.134

Kim, M., & Adlof, L. (2024). Adapting to the future: ChatGPT as a means for supporting constructivist learning environments. *TechTrends, 68*(1), 37–46. https://doi.org/10.1007/s11528-023-00899-x

Koehler, M. J., Mishra, P., & Cain, W. (2013). What is Technological Pedagogical Content Knowledge (TPACK)? *Journal of Education, 193*(3), 13–19. https://doi.org/10.1177/002205741319300303

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. American Educational Research Journal, 32(3), 465–491. https://doi.org/10.3102/00028312032003465

Lei, J. (2009). Digital natives as preservice teachers: What technology preparation is needed? *Journal of Computing in Teacher Education, 25*(3), 87–97. https://doi.org/10.1080/10402454.2009.10784615

Maloy, R. W., Trust, T., Butler, A., & Xu, C. (2021). Critical media literacy and civic learning. EdTech Books. https://edtechbooks. org/mediaandciviclearning

McBean, T., & McBean, T. (2025). Teachers' perspectives and use of artificial intelligence in pedagogy. *International Journal of Social Science Research and Review, 8*(11), 19–34. https://doi.org/10.47814/ijssrr.v8i11.3048

McQuillan, D. (2022). *Resisting AI: An anti-fascist approach to artificial intelligence* (1st ed.). Policy Press. https://doi.org/10.1332/policypress/9781529213492.001.0001

Melese, F. (2025). Review of: Should robots replace teachers? *Curriculum and Teaching Dialogue, 27*(2), 374–378. https://doi.org/10.22488/okstate.25.100635

Meskó, B. (2023). Prompt engineering as an important emerging skill for medical professionals: Tutorial. *Journal of Medical Internet Research, 25*, Article e50638. https://doi.org/10.2196/50638

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing. https://doi.org/10.54675/EWZM9535

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

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

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability, 16*(3), 978. https://doi.org/10.3390/su16030978

Oved, O., & Alt, D. (2025). Teachers' technological pedagogical content knowledge (TPACK) as a precursor to their perceived adopting of educational AI tools for teaching purposes. *Education and Information Technologies, 30*(10), 14095–14121. https://doi.org/10.1007/s10639-025-13371-5

Petko, D., Mishra, P., & Koehler, M. J. (2025). TPACK in context: An updated model. *Computers and Education Open, 8*, Article 100244. https://doi.org/10.1016/j.caeo.2025.100244

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. https://doi.org/10.48550/arXiv.2402.07927

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.3102/0013189X0150020

Van den Berg, G., & Du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences, 13*(10), Article 998. https://doi.org/10.3390/educsci13100998

Velander, J., Taiye, M. A., Otero, N., & Milrad, M. (2024). Artificial intelligence in K-12 education: Eliciting and reflecting on Swedish teachers' understanding of AI and its implications for teaching & learning. *Education and Information Technologies, 29*(4), 4085–4105. https://doi.org/10.1007/s10639-023-11990-4

Wang, K., Ruan, Q., Zhang, X., Fu, C., & Duan, B. (2024). Pre-service teachers' GenAI anxiety, technology self-efficacy, and TPACK: Their structural relations with behavioral intention to design GenAI-assisted teaching. *Behavioral Sciences, 14*(5), Article 373. https://doi.org/10.3390/bs14050373

Willermark, S. (2018). Technological pedagogical and content knowledge: A review of empirical studies published from 2011 to 2016. *Journal of Educational Computing Research, 56*(3), 315–343. https://doi.org/10.1177/0735633117713114