Journal of Technology-Integrated Lessons and Teaching
(2026) Volume 5, Issue 1; DOI: 10.13001/jtilt.v5i1.10317

Artificial Intelligence Literacy Workshop for Graduate Teaching Assistants

Md Mahbubul Amin, Emmanuel Amackson, Samara Santana, and Mohammad Shams Ud Duha, University of Oklahoma

Overview

This training program describes a pilot implementation designed to support Graduate Teaching Assistants (GTAs) in developing competency in the responsible and pedagogical use of generative artificial intelligence (GenAI) in their roles as instructors of record in higher education. Guided by Gagné's (1985) Nine Events of Instruction, Backward Design (Wiggins & McTighe, 2005), and the ADDIE framework (Branch, 2009), the session provides structured instruction in which participants use ChatGPT (5.1) to develop prompt engineering skills and evaluate AI-generated instructional outputs. GTAs construct, test, and iteratively refine discipline-specific prompts using the Role, Task, Context, Format framework (RTCF; Kang et al., 2025) and the AI Assessment Scale (AIAS; Perkins et al., 2024).

Topics: AI Output Evaluation; Generative AI; Prompt Engineering

Time: 2 hours

Materials

- Computers with internet access
- Projector
- Presentation Slides
- Padlet (or another collaborative note tool)
- ChatGPT (5.1)
- Printed Materials

Context-at-a-Glance

Setting
For GTAs in higher education who serve as instructors of record or have other responsibilities in undergraduate courses.

Modality
In-person session

Class Structure
Conducted in a flexible computer lab with individual laptops, shared projection, and movable seating for whole-group modeling, small-group collaboration, and independent prompt experimentation.

Organizational Norms
GTAs had basic familiarity with GenAI tools and responsibility for communicating permissible AI use in their courses. Institutional leadership supported structured professional development in response to emerging AI policy demands (McDonald et al., 2024).

Learner Characteristics
Learners are newly enrolled graduate students at a university in the United States. Learners come from various geographical and academic backgrounds.

Instructor Characteristics
Ph.D. students in learning experience design and technology with expertise in GenAI integration, instructional design, and pedagogical evaluation.

Development Rationale
Many GTAs lack structured preparation to integrate GenAI responsibly into teaching.

Design Framework
Gagné's (1985) Nine Events of Instruction, Backward Design (Wiggins & McTighe, 2005), and ADDIE (Branch, 2009)

SETUP

The learning environment was designed to be well-lit, and free from distractions, with seating arranged to support interaction (e.g., small groups or a semicircle if discussion or collaboration is expected).

The instructors used 15-30 minutes for effective setup and preparation. This included arranging the physical space, testing audiovisual equipment, logging into digital platforms, loading presentation materials, and conducting a brief check to ensure all learners can see, hear, and access required resources.

CONTEXT AND SETTING

Recent developments in GenAI have introduced both instructional opportunities and pedagogical uncertainty within higher education. GTAs, who frequently serve as instructors of record or lead discussion, laboratory, and grading responsibilities in undergraduate courses, are increasingly expected to make instructional and policy decisions regarding GenAI use. However, many GTAs receive limited structured preparation for integrating emerging technologies into their teaching practice. As a result, decisions about prompt construction, output evaluation, and course-level AI policies are often made without shared criteria, formal guidance, or pedagogical grounding (Davis & Lee, 2023; McDonald et al., 2024). At the time of this pilot implementation, the university did not provide formal training for GTAs on responsible GenAI use in classroom contexts, though discussions were underway with university administrators to incorporate such training into future GTA orientation programs. While many online tutorials demonstrate how to use GenAI tools, few professional development resources focus on how instructors can intentionally design prompts and evaluate outputs within authentic teaching contexts. This gap creates risks of inconsistent policy enforcement, overreliance on GenAI outputs, or avoidance of potentially beneficial tools.

In response, we designed a professional development program centered on prompt engineering as instructional decision-making rather than technical manipulation. Prompt engineering refers to the practice of designing structured inputs that guide AI systems toward producing relevant, accurate, and pedagogically appropriate outputs. In higher education contexts, prompt engineering functions as a mechanism through which users craft prompts to generate more desirable outcomes, such as improved accuracy, relevance, and instructional applicability (Lee & Palmer, 2025). The conceptual foundation of the program positions prompt construction as a cognitive and pedagogical act in which instructors externalize instructional intent, learning objectives, and constraints to guide AI-generated outputs. The full program was structured as a four-unit sequence spanning foundational AI literacy, ethics and institutional policy, applied prompt engineering, and course-level policy reflection. However, only the Prompt Engineering and Application unit was implemented in a pilot format to evaluate instructional effectiveness, feasibility, and alignment with intended learning outcomes. The remaining units were fully developed but remain conceptual for future full-scale implementation. Table 1 summarizes the structure, modality, and implementation status of each unit.

Table 1

Overview of Program Structure and Modality

Unit	Title	Modality	Status
1	AI Literacy and Foundations	Online Asynchronous	Designed conceptual
2	Ethics and Institutional Policy	Online Asynchronous	Designed conceptual
3	Prompt Engineering and Application	In-Person	Implemented pilot
4	Course-Specific Policy Design and Reflection	In-Person	Designed conceptual

We developed the training program using complementary instructional design frameworks. Backward Design (Wiggins & McTighe, 2005) informed our identification of desired transfer outcomes because the program aimed to develop transferable instructional decision-making skills rather than isolated technical tasks. This approach required prioritizing learning outcomes and assessment evidence before planning instructional activities. Learning objectives therefore focused on enabling GTAs to construct aligned prompts, evaluate outputs using defined criteria, and exercise instructional judgment when using GenAI. We used ADDIE (Branch, 2009) to guide the broader development process of the program, allowing needs assessment findings, iterative refinement, and pilot evaluation to inform design and implementation decisions. Within this structure, we organized the instructional sequence of the Prompt Engineering and Application unit using Gagné's (1985) Nine Events of Instruction, which structured the progression from attention and modeling to guided practice, feedback, and transfer during the session. In this design, Backward Design structured the identification of learning outcomes and assessment evidence, ADDIE guided the overall program development process, and Gagné's (1985) framework organized the pedagogical flow of the implemented session, allowing the frameworks to function as complementary layers rather than competing models.

Needs assessment data indicated that participating GTAs represented diverse disciplinary backgrounds, including humanities, social sciences, STEM, and professional programs, and varied widely in prior experience with GenAI tools. Approximately two-thirds reported experimenting with tools such as ChatGPT (5.1) for instructional tasks, including lesson planning, discussion question generation, and feedback preparation, while others reported minimal prior use. Despite this exposure, about two-thirds expressed uncertainty about how to judge whether AI-generated outputs were pedagogically appropriate, accurate, or ethically responsible for classroom use. Interviews with subject matter experts (SMEs), including faculty mentors and instructional designers, reinforced these findings, noting that many GTAs relied on ad hoc prompting strategies and lacked shared criteria for evaluating AI outputs. These findings informed the pilot session design by emphasizing explicit modeling of prompt construction, rubric-guided evaluation using the AI Assessment Scale (AIAS; adapted from Perkins et al., 2024), and iterative prompt revision using the RTCF framework (adapted from Kang et al., 2025).

The pilot implementation occurred in a technology-enhanced learning lab designed to support active, discussion-based engagement. The physical space included movable seating, individual laptops, shared projection capability, and reliable internet access. This configuration allowed rapid transitions between instructor modeling, paired analysis, independent prompt construction, and whole-group reflection. Because the learning goals required observable performance and revision cycles, the physical and digital environment was intentionally designed to minimize friction and maximize interaction.

The instructional approach emphasized meaning making, structured practice, and iterative feedback rather than passive exposure to AI tools. Participants engaged in real-time prompt testing using ChatGPT (5.1), evaluated outputs using the AI Assessment Scale (adapted from Perkins et al., 2024), and revised prompts based on peer and instructor feedback. The setting supported collaborative problem solving, allowing GTAs to compare outputs across disciplines and reflect on how subtle changes in role specification, contextual detail, or constraints affected instructional quality.

Importantly, although the broader program includes units focused on AI literacy and institutional policy development, the pilot session concentrated specifically on prompt engineering and evaluation competencies. This focus ensured depth of skill development within the available two-hour timeframe and allowed for targeted evaluation of procedural learning outcomes. Future implementations will integrate the conceptual units to provide a more comprehensive professional development sequence.

By situating prompt engineering within authentic teaching responsibilities and grounding design decisions in established instructional frameworks, this session sought to move beyond general AI awareness toward disciplined, pedagogically aligned practice.

Learning Representation

Unit 3 was the only unit of the training program fully implemented during the pilot and served as the performance-centered core of the instructional design (see Table 1). We implemented this two-hour, in-person session to develop GTAs' procedural and conceptual competence in prompt engineering and in the critical evaluation of GenAI-generated instructional outputs. During the session, participants engaged in instructor modeling, structured practice, peer evaluation, and iterative revision of discipline-specific prompts connected to their own teaching contexts.

The design of the session followed principles reflected in Gagné's (1985) Nine Events of Instruction, emphasizing attention, modeling, guided practice, feedback, and opportunities for transfer. Table 2 lists the nine instructional events that guided the pedagogical structure of the session.

Table 2
Gagné's (1985) Nine Events of Instruction

Event	Gagné's Instructional Event
1	Gain Attention
2	Inform Learners of Objectives
3	Stimulate Recall of Prior Knowledge
4	Present the Content
5	Provide Learning Guidance
6	Elicit Performance (Practice)
7	Provide Feedback
8	Assess Performance
9	Enhance Retention and Transfer

Presentation of content and materials were delivered through Microsoft PowerPoint to structure the session, highlight key prompt-engineering strategies, and visually model examples of effective and ineffective AI prompts in teaching contexts. Participants also used Padlet as a shared collaborative space to post prompts, review peers' outputs, and provide feedback during guided practice activities.

Learning Objectives

The learning objectives for Unit 3 are derived directly from prior needs assessment, task analysis, and SME interviews, which identified prompt engineering and AI output evaluation as critical competency gaps for GTAs.

The objectives of this unit were (see Presentation Slide 2):

By the end of Unit 3, participants will be able to:

1. Construct discipline-specific AI prompts using a structured prompt framework that specifies role, task, context, and constraints.
2. Evaluate AI-generated instructional outputs for accuracy, bias, pedagogical alignment, and transparency using the AI Assessment Scale (AIAS)
3. Iteratively revise prompts to improve output quality and reduce ethical or instructional risks.
4. Articulate how prompt engineering supports instructional judgment rather than replacing instructor decision-making.

These objectives reflect higher-order cognitive and procedural outcomes and align with cognitive strategy teaching recommendations (Morrison et al., 2019).

Session Structure and Flow

Unit 3 was implemented using Gagné's (1985) Nine Events of Instruction to ensure that learners were cognitively and motivationally prepared before engaging in independent performance and transfer. The session structure also reflected principles of Backward Design, in which learning objectives and assessment evidence were defined prior to activity sequencing, ensuring that each instructional phase directly supported targeted performance outcomes. In addition, the broader development and refinement of the unit followed the iterative logic of the ADDIE framework, guiding analysis, design, development, implementation, and evaluation of decisions throughout the pilot process.

We intentionally structured the session so that attention and relevance were established early; procedural knowledge was modeled and practiced with support, and opportunities for feedback, reflection, and application were embedded throughout the lesson. This sequencing allowed GTAs to move from guided observation to independent prompt construction and evaluation within a single instructional episode. Table 3 summarizes the structure and timing of the implemented session and illustrates how each phase aligned with Gagné's (1985) instructional events.

Table 3
Unit 3 Session Flow Aligned with Gagné's Events

Phase	Time (Min)	Gagné Event(s)
1: Introduction & Orientation	0-8	1-2
2: Modeling Prompt Engineering	8-20	3-5
3: Guided Diagnosis Practice	20-40	5-6
4: Independent Prompt Construction	40-55	6-7
5: AI Output Evaluation (AIAS)	60-80	6-8
6: Iterative Revision	80-90	7-9
7: Pedagogical Integration	90-110	9
8: Closing & Exit Ticket	110-120	2 & 9

Phase 1: Introduction and Orientation

Before beginning the session, each participant was given a copy of the Printed Materials (DOC). After this, the session began by situating prompt engineering within GTAs' everyday teaching responsibilities, explicitly connecting the lesson to common instructional tasks such as lesson planning, assignment design, and student feedback. To activate prior knowledge and surface variation in participants' prior experience with AI tools, we displayed two opening questions on the presentation slides (see Presentation Slide 3) and invited participants to respond through brief whole-group discussion.

- How many of you have used AI to plan lessons, generate assignment ideas, or write feedback?
- How confident were you in the quality or appropriateness of the AI output?

This brief whole-group interaction helped normalize varied levels of experience, reduced affective barriers, and aligned with Gagné's (1985) first event by gaining learners' attention and establishing relevance. We then presented the session objectives and explicitly framed AI as a supportive instructional tool rather than an autonomous decision-maker, a positioning that directly addressed ethical concerns identified through prior SME interviews and learner analysis.

Phase 2: Modeling Prompt Engineering

We introduced prompt engineering as a cognitive and instructional design practice that required instructors to externalize their pedagogical intent in ways that generative AI systems could interpret and act upon. We emphasized that effective learning of prompt design requires learners to first observe expert modeling before engaging in independent performance. We explicitly framed prompt engineering not as a technical shortcut, but as an extension of instructional decision-making grounded in learning objectives, learner characteristics, and ethical considerations.

To support this understanding, we presented a structured prompt framework (adapted from Kang et al., 2025) consisting of four components, acronymized as RTCF (see Presentation Slide 4):

- **Role:** the instructional identity the AI should assume (e.g., teaching assistant, tutor, rubric designer).
- **Task:** the specific instructional action to be completed (e.g., generating discussion questions, drafting feedback, creating assessment items).
- **Context:** relevant information about the course, learner level, learning goals, and known misconceptions.
- **Format/Constraints:** expectations for output structure, tone, length, and ethical guardrails (e.g., avoiding answer generation or bias reinforcement).

We explained each component using concrete, teaching-related examples to ensure accessibility across disciplinary backgrounds. We then conducted a live demonstration using two prompts addressing the same instructional task: generating discussion questions for a class on academic integrity, one underspecified, and one deliberately structured using the RTCF framework (see Presentation Slide 5). The weak prompt presented to participants was:

- "Create discussion questions about academic integrity."

In contrast, the structured prompt followed the RTCF framework:

- "You are a graduate teaching assistant preparing a class discussion for a first-year undergraduate course. Generate three open-ended discussion questions about academic integrity and AI use in assignments. The questions should encourage critical thinking, include one short classroom scenario, and avoid yes/no responses."
- **R**: You are a graduate teaching assistant
- **T**: Generate three open-ended discussion questions about academic integrity and AI use in assignments
- **C**: preparing a class discussion for a first-year undergraduate course
- **F**: The questions should encourage critical thinking, include one short classroom scenario, and avoid yes/no responses."

Both prompts were run in real time using ChatGPT (5.1), and the resulting outputs were displayed and compared publicly. During this demonstration, we employed a think-aloud strategy to articulate our instructional reasoning, making the invisible cognitive work of prompt construction explicit and showing how variations in role specification, contextual detail, and constraints influenced the quality and appropriateness of AI-generated outputs. This modeling phase aligned with Gagné's (1985) Events 4 and 5 by presenting essential content and providing learning guidance before learners attempted the procedure independently.

Phase 3: Guided Diagnosis Practice

Following the modeling phase, participants engaged in a guided practice activity designed to surface common prompt design errors while keeping cognitive load manageable. Working in pairs, GTAs analyzed pre-written weak prompts drawn from familiar instructional contexts (e.g., "Summarize the Civil Rights Movement;" "Give feedback on this essay") to identify shortcomings in clarity, contextualization, and ethical framing. Pairs then diagnosed issues such as ambiguity, missing instructional constraints, or lack of guidance regarding appropriate AI assistance and collaboratively rewrote the prompts using the structured prompt framework (see Printed Materials DOC, Common Pitfalls section, for common issues).

During this activity, we circulated among groups to provide immediate formative feedback, pose clarifying questions, and redirect attention to key elements of effective prompt design when needed. Placing this guided practice immediately after the modeling phase supported early skill acquisition by allowing GTAs to apply newly introduced concepts in a low-stakes environment.

Phase 4: Independent Prompt Construction

Participants then independently constructed two prompts aligned with their own teaching responsibilities (see Presentation Slide 6). The first prompt focused on an instructional design task (e.g., lesson planning, quiz construction, or discussion question development), while the second focused on feedback or tutoring, explicitly incorporating ethical boundaries such as limiting answer generation or requiring guided questioning. To support instructional alignment, we asked participants to begin by pasting a relevant course learning objective into a shared Padlet workspace before drafting each prompt. This shared digital board allowed participants to post their prompts, observe peers' approaches across disciplines, and comment on each other's work during the activity.

During this phase, we circulated throughout the room to offer targeted feedback, prompting participants to clarify instructional intent, refine contextual details, and strengthen ethical framing where needed. For example, when a participant posted a prompt such as "Generate quiz questions about photosynthesis," we asked follow-up questions such as "What level of students are you teaching?" or "What type of thinking do you want students to demonstrate?" to help them revise the prompt using the structured RTCF framework. This activity required learners to apply the prompt framework independently, corresponding to Gagné's (1985) Event 6 (eliciting performance), while ongoing instructor feedback supported Event 7 (providing feedback).

Phase 5: AI Output Evaluation (AIAS)

Participants then generated AI outputs using their drafted prompts and exchanged outputs with a peer for evaluation. Prompts and outputs were posted on Padlet, allowing peers to review examples across different teaching contexts and provide written feedback. Each peer evaluated the AI-generated output using the AI Assessment Scale (AIAS), adapted from Perkins et al. (2024; see Presentation Slides 7-8). We introduced the AIAS as a criterion-referenced tool for judging both ethical and pedagogical quality, emphasizing its relevance to real instructional decision-making. The scale operationalized evaluation across four dimensions: accuracy, bias/fairness, pedagogical alignment, and transparency. Table 4 summarizes the tool.

Table 4
AI Output Evaluation Criteria (AIAS)

Dimension	Description
Accuracy	Factual correctness and conceptual soundness
Bias/Fairness	Inclusive language and avoidance of stereotypes
Pedagogical Alignment	Alignment with stated learning objectives
Transparency	Disclosure of limitations and uncertainty

Peers used the rubric to comment on specific strengths and weaknesses in the generated outputs. For instance, one peer noted that an AI-generated discussion question aligned well with the learning objective but lacked contextual constraints, recommending that the prompts specify the course level and desired format of the response. Peers then suggested at least one revision to the prompt that could improve the output on one or more AIAS dimensions. This structured peer-evaluation activity supported criterion-referenced judgment, reinforced ethical awareness, and mirrored the types of evaluative decisions GTAs routinely make when selecting or revising instructional materials (Events 6-8).

Phases 6-7: Iterative Revision & Pedagogical Integration

Based on AIAS feedback, participants revised and reran their prompts to test the effects of their changes on AI-generated outputs. Revised prompts and outputs were reposted on Padlet, allowing participants to compare initial and revised versions and observe how modifications in role specification, contextual detail, or output constraints affected response quality. Participants were asked to note whether their revisions improved the output on at least one AIAS dimension (e.g., accuracy, pedagogical alignment, or transparency).

This rapid cycle of revision and re-execution reinforced prompt engineering as an iterative design process rather than a static skill. By immediately applying peer and instructor feedback to refine their work, participants strengthened retention and supported transfer, aligning with Gagné's (1985) Event 9 (Enhancing Retention and Transfer) through learning by application and reflection.

Phase 8: Closing & Exit Ticket

We concluded the session with a brief exit-ticket reflection designed to promote transfer and integration of prompt engineering into participants' teaching contexts. GTAs responded in writing to two prompts (see Presentation Slide 9):

1. One concrete change I will make to a lesson, assignment, or feedback practice using AI prompting ....
2. One way I will use prompt engineering to support my instructional judgment while setting clear boundaries for AI use in my course: ....

These prompts asked participants to connect prompt engineering to immediate instructional decisions while reflecting on responsible AI use in their teaching practice. We closed the session by reinforcing key principles and encouraging continued refinement of prompts in future courses (see Presentation Slide 10). Representative responses reflected this connection. One GTA wrote, "I will use structured prompts to generate draft discussion questions and then revise them so they match my learning objectives;" while another noted, "AI can help with brainstorming or outlining feedback, but final grading decisions should always come from me."

Example of Prompt Development and Revision

The following example illustrates how participants applied the prompt framework during the activity, evaluated AI-generated outputs, and refined prompts based on peer feedback shared through Padlet. The example reflects typical feedback patterns observed during the session, and participant comments are presented anonymously.

Task: Generate discussion questions for a first-year undergraduate course addressing academic integrity and AI use.

Prompt: "You are a graduate teaching assistant preparing a class discussion for a first-year undergraduate course. Generate three open-ended discussion questions about academic integrity and AI use in assignments."

AI Output: What does academic integrity mean in the context of using AI tools for coursework?

Can using AI tools like ChatGPT ever violate academic integrity policies? Why or why not?

What responsibilities do students have when using AI tools to assist with their assignments?

Peer Feedback: Peers commented on the initial posts, noting: "These questions are good, but they feel a bit general. I think you can include a short scenario so students can evaluate a real situation." Another peer added, "These look good to me, but I'd probably add a specific example of AI use in an assignment. It might spark more critical reflections."

2nd Prompt (Building on the First Prompt): "Add a short classroom scenario involving a student using AI to help write part of an essay, and make the questions encourage critical analysis."

Revised AI Output: Scenario: A student uses an AI tool to generate a paragraph for an essay and then edits the paragraph before submitting the assignment without mentioning the AI tool.

1. In this scenario, do you think the student's use of AI violates academic integrity policies? Why or why not?
2. What responsibilities do students have when deciding whether and how to use AI tools in their academic work?
3. How might instructors clarify expectations about AI use so that situations like this are easier for students to navigate?

By this stage of the activity, GTAs recognized that prompts could be iteratively refined by adding additional details or constraints to guide the AI toward more relevant and instructionally aligned outputs.

Assessment

Assessment in this session was formative and performance-based, focusing on participants' ability to apply prompt engineering principles in authentic teaching contexts. We collected multiple forms of evidence to capture both procedural skill development and transfer-oriented thinking. Evidence of learning included two finalized discipline-specific prompts, AI-generated outputs evaluated using the AI Assessment Scale (AIAS), written peer feedback artifacts, and a brief exit-ticket reflection focused on future instructional application. Representative examples of initial prompts, revised prompts, AIAS evaluations, and exit-ticket responses are provided in the previous sections to illustrate how participants applied the prompt framework and refined their work during the session.

Overall, participant artifacts indicated that all the participating GTAs were able to construct structured prompts aligned with instructional goals and improve output quality through iterative revision. Peer evaluations using the AIAS frequently identified opportunities to strengthen contextual detail, clarify instructional constraints, and improve pedagogical alignment. Exit-ticket reflections further suggested that participants were able to articulate how prompt engineering could support instructional planning while maintaining instructor judgment and ethical boundaries in AI use.

Table 5 summarizes the alignment between instructional objectives, assessment tasks, and evidence collected during the session.

Table 5
Assessment Alignment for Unit 3

Instructional Objective	Assessment Task	Evidence
Construct effective, discipline-specific AI prompts	Prompt Design Task (2 prompts)	Finalized prompts
Evaluate AI-generated outputs critically and ethically	AI Output Review (AIAS)	Annotated AI outputs
Revise prompts based on feedback	Prompt Revision	Revised prompts and outputs
Articulate prompt engineering to teaching practice	Exit-Ticket Reflection	Written reflection

Evaluation

Formative evaluation during Unit 3 was conducted through a combination of instructor feedback, peer guidance, and independent practice. We facilitated pedagogically relevant, real-world reflective practice by using performance-based assessment tools, including finalized prompts, written reflections, and revised AI-generated outputs. Each performance artifact was evaluated against its corresponding aligned rubric, allowing us to monitor participants' procedural accuracy, ethical reasoning, and instructional alignment as learning unfolded.

In addition to evaluating participant performance, we asked the instructors involved in the pilot to provide feedback on instructional effectiveness, feasibility, and implementation considerations. This feedback focused on issues such as clarity of instructions, pacing of activities, facilitation strategies, and the practicality of activities within time constraints. We used these insights to inform subsequent lesson planning and to identify areas requiring additional scaffolding or adjustment in future iterations.

Summative evaluation was informed by cumulative assessment data aligned with the relevant rubrics, as well as thematic analysis of open-ended survey responses addressing both participant learning and instructional effectiveness. Aggregated results were used to determine the proportion of learning objectives met across participants. Following Morrison et al.'s (2019) guidance for evaluating skill-based instructional programs, we examined outcomes relative to the benchmark that programs in which 90% of learners achieve 90% of objectives may be considered highly effective. Assessment results from the pilot were interpreted in relation to this criterion to evaluate overall instructional effectiveness.

Evaluation data from the pilot implementation suggested that the instructional objectives of Unit 3 were largely met. Review of participant artifacts indicated that most GTAs were able to construct two discipline-specific prompts that demonstrated clear instructional intent, appropriate contextualization, and explicit ethical constraints. Comparisons of AI-generated outputs before and after revision showed observable improvement in output quality for many participants, particularly in terms of pedagogical alignment and transparency as defined by the AIAS criteria. Exit-ticket reflections further indicated evidence of transfer, as participants consistently identified specific instructional uses for AI prompting and articulated ethical guardrails they intended to communicate to students.

Taken together, these findings provided preliminary evidence that the instructional design supported procedural skill development and ethical reasoning related to prompt engineering within the pilot context.

Critical Reflection

We implemented the lesson once with a small cohort of GTAs who demonstrated varying levels of familiarity with GenAI tools. Review of participant artifacts and exit-ticket responses suggested that the session generally met its intended objectives: participants were able to construct structured prompts, evaluate AI-generated outputs, and revise prompts for instructional tasks. However, the pilot implementation also revealed several design limitations that would need to be addressed in future iterations.

One limitation we encountered involved differences in participants' prior experience with GenAI. For example, during the independent prompt construction activity, several GTAs quickly generated well-structured prompts and proceeded to experiment with refinements, while others needed additional clarification about how to specify roles, contextual information, or output constraints within their prompts. As a result, pacing during the activity became uneven, with some participants waiting for others to complete the initial prompt drafting stage. In future implementations, we would address this issue by incorporating tiered activities, such as providing scaffolded prompt templates for beginners while offering optional extension tasks that allow more experienced participants to explore advanced prompt strategies or disciplinary applications.

A second limitation was related to time allocation during the session. Although the two-hour structure allowed us to model prompt construction, facilitate guided practice, and conduct peer evaluation using the AIAS, the iterative revision phase was shorter than originally intended. Several participants indicated during informal discussion that they would have liked additional time to test multiple prompt variations and compare AI outputs. In future versions of the workshop, we would extend Phase 6: Iterative Revision or assign an additional prompt-refinement activity that participants could complete after the session.

Technology logistics also influenced the session design. Because participants generated prompts and evaluated outputs in real time, reliable access to laptops and internet connectivity was essential. During the pilot, we observed that even brief connectivity delays slowed transitions between activities. If the workshop were implemented in contexts where individual devices are unavailable, we would adapt the activity by conducting instructor-led demonstrations using a shared display and facilitating group analysis of prompts and outputs.

Another important constraint was that this session was implemented as a standalone unit rather than as part of the full four-unit training sequence originally designed for the program. Because participants had not completed the earlier units on AI literacy and institutional policy, we needed to spend additional time clarifying foundational concepts that were intended to be introduced earlier in the sequence. In future implementations, we would integrate the units as originally designed so that participants enter the prompt engineering session with a shared understanding of GenAI capabilities, limitations, and institutional expectations.

Finally, while the pilot evaluation provided evidence of immediate skill development and participants reflected on how they would integrate the AI prompts into their pedagogy, we were not able to examine how participants actually applied these strategies in their teaching after the workshop (Phase 7: Pedagogical Integration). Future implementations could incorporate follow-up surveys, teaching artifact analysis, or instructor reflections to better understand how prompt engineering skills transfer into authentic instructional practice over time.

Taken together, these reflections suggest that the workshop structure is adaptable across teaching contexts but would benefit from stronger scaffolding, improved time allocation for revision activities, and full integration within the broader training sequence.

References

An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education, 22*(1), Article 10. https://doi.org/10.1186/s41239-025-00507-3

Branch, R. M. (2009). *Instructional design: The ADDIE approach.* Springer.

Davis, R. O., & Lee, Y. J. (2023). Prompt: ChatGPT, create my course, please! *Education Sciences, 14(1*), Article 24. https://doi.org/10.3390/educsci14010024

Gagné, R. M. (1985). *The conditions of learning* (4th ed.). Holt, Rinehart and Winston.

Kang, W., Zhang, L., Peng, X., Zhang, H., Li, A., Wang, M., Huang, J., Tian, F., & Dai, G. (2025). TutorCraftEase: Enhancing pedagogical question creation with large language models. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems,* Article 1076, 1-22. https://doi.org/10.1145/3706598.3713731

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education, 22*(1), Article 7. https://doi.org/10.1186/s41239-025-00503-7

McDonald, N., Johri, A., Ali, A., & Hingle, A. (2024). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *arXiv.* https://doi.org/10.48550/arXiv.2402.01659

Morrison, G. R., Ross, S. M., Morrison, J. R., & Kalman, H. K. (2019). *Designing effective instruction* (8th ed.). Wiley.

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of Generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(06). https://doi.org/10.53761/q3azde36

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (2nd ed.). ASCD.

About the Authors

Md Mahbubul Amin is a Ph.D. student in Learning Experience Design and Technology at the University of Oklahoma and works as a Graduate Research Assistant focusing on instructional design and AI literacy initiatives. His professional experience spans instructional design, language education, and digital learning development across higher education and professional training contexts. His research and professional interests include AI-enhanced learning, instructional design for higher education, language learning technologies, and bridging research and practice to support educators through practical, accessible learning solutions.

Emmanuel Amackson is a Ph.D. student in Learning Experience Design and Technology at The University of Oklahoma. His research focuses on the design and evaluation of technology-enhanced learning environments, with interests in AI-supported training, adaptive learning systems, and game-based learning. Emmanuel is interested in exploring how emerging technologies can be used to personalize learning experiences, foster engagement, and support

meaningful learning across educational and professional contexts.

Samara Santana is a master's student in the Science of Psychology, Research, and Data in Education at the University of Oklahoma and works as a research technician alongside the Center of Child Abuse and Neglect (CCAN) at the University of Oklahoma Health Science Center. As a research technician, her research agenda centers on generating evidence-based level-of-need recommendations that promote youth permanency. This has informed her perspective on how educational contexts can cultivate community and meet the needs of diverse learners through research interests in Universal Design for Learning, visual literacy, and instructional message design that integrates both theory and practice.

Mohammad Shams Ud Duha is an Assistant Professor of Learning Experience Design and Technology in the Department of Educational Psychology at the University of Oklahoma. He earned his Ph.D. in Learning Design and Technology from Purdue University. His research interests include online learning environments, social media in education, open education, Universal Design for Learning, and microlearning. His work has been published in reputed journals such as Online Learning, Distance Education, Journal of Computing in Higher Education, and TechTrends.

Sharing & Modification Permissions