

AI TEACHING ASSISTANTS FOR TEACHER EDUCATION: A MIXED-METHOD STUDY*

Noam Topelberg, & Esther Isman
Faculty of Education, Bar-Ilan University (Israel)

Abstract

In the past two years, large language models have become an integral part of learning and teaching in schools and higher education. Given the rapid integration of artificial intelligence into everyday life, and into educational systems in particular, a comprehensive, multidisciplinary, and systematic approach is imperative to ensure both high-quality teaching and the development of essential learner competencies can be achieved effectively. Our team, proficient in diverse fields in both regular and special education, participated in a pilot program initiated by Bar-Ilan University. In this program, AI-based teaching assistants were developed for selected courses. Following this, we developed an AI-based teaching assistant specializing in planning, writing, and providing feedback on lesson plans and learning processes, which was then adapted into course-specific versions. Throughout the 2025 academic year, we implemented these assistants into our teacher education classes and conducted a mixed-method study using qualitative and quantitative questionnaires to examine students' perceived mastery and self-assessment of lesson planning skills. Participants included students in general education ($n = 67$), special education ($n = 42$), and a general education control group ($n = 57$). Findings indicate high student satisfaction and perceived support for clarifying assignment requirements and providing rapid, personalized feedback. In addition, there is qualitative evidence for the development of prompt literacy and critical monitoring. However, the analysis also highlights the importance of explicit instruction in prompt literacy and critical evaluation of AI outputs. These findings contribute to the emerging literature on integrating AI-based teaching assistants in teacher education.

Keywords: *AI teaching assistants, teacher education, lesson planning, self-regulated learning, formative feedback.*

1. Introduction

With the advent of large language models (LLMs) in late 2022, higher education institutions have begun experimenting widely with AI-based teaching assistants that generate text, simulate dialogue and provide feedback in response to student requests. These developments raise a central question for teacher education: how can AI tools be integrated in ways that support, rather than undermine, core pedagogical practices and professional judgement. Recent work suggests that custom AI assistants tailored to specific curricula, institutional policies and pedagogical goals can help align AI use with the values and standards of particular teacher education programs (Almasre, 2024; Masters et al., 2024).

Studies on educational chatbots highlight both opportunities and risks. Dedicated assistants can provide personalized guidance, timely feedback and accessible support, yet they also introduce challenges related to overdependence, academic integrity and fabricated outputs (Chang et al., 2023; Kuhail et al., 2023). Recent findings indicate that generative AI may reduce perceived cognitive effort and lower users' tendency to scrutinize AI generated responses (Lee et al., 2025). For teacher education this dual effect is especially significant: AI can restructure routine preparation and assessment tasks, but it may also weaken deliberate planning, critical reflection and adaptive decision-making. In this context, custom AI teaching assistants that are tightly coupled to local pedagogical frameworks appear particularly promising, since they can draw on course syllabuses, academic readings and policy documents to provide contextually appropriate feedback on lesson planning and related tasks (Almasre, 2024; Chang et al., 2023).

Our study is situated within a teaching training program at Bar-Ilan University's Faculty of Education, where careful planning of teaching and learning processes, and especially lesson planning, is regarded as a core professional skill across diverse educational settings. We developed an integrated knowledge base on lesson plan design that brings together generic principles, disciplinary perspectives and

* The authors wish to thank Ms. Hagar Pinus, for her insightful comments and data analysis.

special education considerations. Two complementary tools were created using this shared corpus: A digital learning module that introduces and refreshes key concepts in lesson planning, and a family of custom AI chatbots. The latter includes a generic assistant for writing lesson plans, course specific assistants that combine course content with the shared lesson planning framework, and discipline specific bots that connect national curricula and teaching materials with the pedagogical model. These tools are intended to support students both in their coursework and in their school-based clinical practice.

Empirical work on such tightly coupled custom-designed AI tools in teacher education is still limited. Most existing studies focus on generic chatbots in higher education or on AI tools in specific subject domains (Almasre, 2024; Kuhail et al., 2023). There is relatively little evidence on how program-aligned assistants influence pre-service teachers' use of AI in lesson planning and how they experience such tools in their day-to-day practicum. Drawing on Zimmerman's framework for self-regulated learning, which conceptualizes learning as cycles of forethought, performance and self-reflection, this study examines how a custom assistant can support planning, monitoring and evaluation processes in the specific context of lesson planning (Zimmerman, 2000; Chang et al., 2023).

2. Background and related work

Early AI educational technologies focused on automated marking, plagiarism detection and basic tutoring systems with limited interactivity. Recent advances in large language models have enabled chatbots that generate context-sensitive responses, adapt to learners' inputs and provide ongoing formative feedback (Chang et al., 2023; Kuhail et al., 2023). These developments create new opportunities for guidance at scale, but also place fresh demands on teacher education, as future educators must learn to use such tools responsibly and critically.

Systematic reviews of educational chatbots highlight this inherent tension. Kuhail et al. (2023) show that chatbots can foster engagement, provide timely guidance and support large cohorts, yet they also raise concerns about accuracy, trust and the lack of robust evaluation frameworks. Chang et al. (2023) similarly emphasize that risks such as "hallucination", the production of plausible but incorrect information, require safeguards, especially in domains where precision is essential. Recent survey findings further suggest that generative AI may reduce perceived cognitive effort and lower users' tendency to scrutinize AI generated responses (Lee et al., 2025), reinforcing the need for deliberate pedagogical design.

Theoretical work on AI enhanced learning frequently draws on self-regulated learning (SRL) perspectives. Zimmerman (2000) conceptualizes learners as proactive agents who plan, monitor and reflect on their work. When AI assistants are designed around SRL principles, they can prompt students to clarify goals, track progress and adjust strategies, embedding planning, monitoring and evaluation cycles within a digital environment (Chang et al., 2023). From this perspective, the question is not only whether AI tools provide correct answers, but how they shape learners' habits of preparation, inquiry and reflection.

Within this broader field, custom AI tools have emerged as a particularly promising approach. Rather than relying on general purpose models, custom assistants are configured on domain specific corpora, such as course syllabuses, scholarly readings and institutional policy documents, in order to generate contextually appropriate, pedagogically aligned responses (Almasre, 2024). In a higher education design course, for example, a custom assistant was used to assess typography students' work, offering standardized feedback based on predefined criteria while remaining sensitive to course specific expectations (Almasre, 2024). Recent literature suggests that such tools can support personalized learning, consistent feedback, scalability and multilingual access for diverse student populations (Aithal & Aithal, 2024; Chang et al., 2023; Masters et al., 2024).

At the same time, scholars caution against uncritical adoption of generative AI in professional education. If lesson plans and pedagogical resources are produced too easily, there is a risk that preservice teachers will bypass essential processes of analysis, justification and reflective decision making. Authors therefore argue for "pedagogical alignment", in which AI tools are integrated into existing teaching frameworks as co facilitators rather than surrogates, and are accompanied by explicit training in critical evaluation of AI outputs (Lee et al., 2025; Masters et al., 2024). From a "glocal" perspective, global AI capabilities must also be tuned to local curricular standards, accreditation requirements and pedagogical traditions (Aithal & Aithal, 2024). In multilingual and multicultural educational settings, this implies careful curation of the knowledge base, ongoing verification of content and the incorporation of culturally relevant examples into chatbot interactions.

Against this backdrop, the present study examines a family of faculty designed custom chatbots that are tightly coupled to a shared lesson planning framework within a teacher education program. While previous work has often focused on generic chatbots or single course implementations, there is still limited empirical evidence on how program aligned assistants are actually used by students in a teacher training program, and how they experience the support these tools provide for both the completion of course requirements and aiding students taking part in practicum.

2.1. Research objectives

Our study seeks to address the following objectives:

1. To examine the contribution of integrating a chatbot as a teaching assistant in academic courses for pre-service teachers.
2. To investigate whether interaction with the chatbot assists pre-service teachers in improving their lesson planning skills through processes of receiving and implementing feedback
3. To identify key factors in the effective use of dedicated chatbots for the development of required skills in teacher education.

3. Method

3.1. Context and participants

In 2025, we enrolled 166 undergraduate students in courses focused on lesson planning and classroom practice. Participants were divided into three groups: Two experimental groups, one in general education and one in special education, whose coursework included a dedicated artificial intelligence teaching assistant, and a control group of general education students who did not use the AI assistant.

Using a mixed method approach, we combined quantitative survey data with qualitative accounts of students' experiences. A structured Likert type questionnaire was administered to three groups of pre-service teachers to assess their use of AI tools in lesson planning and their perceptions of the contribution of a dedicated AI teaching assistant. Students of the two experimental groups, who had direct access to the AI-based teaching assistant, also responded to two open-ended questions regarding the use of an AI teaching assistant in the course. The qualitative responses were analyzed thematically and used to complement the statistical findings, allowing us to interpret patterns of AI use and perceived support in light of students' own descriptions of when the assistant was helpful, when it fell short and what kinds of prompting and pedagogical conditions they regarded as necessary for effective use.

A total of 67 students were included in the experimental general education group, of which 48 students (71.6%) were female, and 19 students (28.4%) were male. In this group, 41 students (61.2%) were not teaching at the time of the study, 12 (17.9%) were teaching only as part of a school-based practicum, and 14 (20.9%) had at least one year of teaching experience. 51 students (76.1%) indicated no prior use of artificial intelligence before their studies at university, while 16 (23.9%) indicated previous use of artificial intelligence. An additional experimental special-education group comprised 42 students (25.3% of the sample), including 41 females (97.6%) and one male (2.4%). Most of these students ($n=39$, 93%) taught as part of their practicum during the study, and three students (7.2%) had teaching experience at least one year prior. There were 26 students (61.9%) in this group who had no prior experience with artificial intelligence tools, whereas 16 students (38.1%) reported experiencing them previously.

There were 57 general-education students in the control group (34.3% of the sample), including 46 women (80.7%) and 11 men (19.3%). Most ($n=39$, 68.4%) of the participants were not teaching at the time of the study, 8 (14.0%) were teaching as part of their practicum programs, and 10 (17.5%) had been teaching at least one year. Most of the students in this group (59.6%) had no prior experience with AI tools, whereas 23 (40.4%) reported having used AI tools before their studies at the university.

Across the full sample, 119 students (71.7%) were aged between 20 and 30, and 47 (28.3%) were aged 31 or above. Background variables - gender, age group, teaching status, and prior AI experience - were examined in relation to the dependent measures and included as covariates where statistically significant associations were found.

3.2. Instruments

Data for this study were collected using a self-constructed questionnaire developed specifically for the purposes of this research. The questionnaire included three sections:

Background: Participants answered background questions regarding gender, age group, teaching status (not teaching; practicum only; at least one year of teaching), and prior AI experience. Using these variables, we were able to characterize the sample and examine confounding factors.

Sense of Skill mastery: Participants answered questions assessing the extent to which they sense mastery in ten aspects of developing lesson plans, including writing goals and operational objectives, lesson structure planning, understanding levels of thinking according to Bloom's taxonomy, and identifying learning styles to enhance learning and more. Each item was graded on a five-point Likert scale. The internal consistency on this scale was high (Cronbach's $\alpha = .891$).

To evaluate the perceived contribution of AI teaching assistants to skill mastery, the two experimental groups rated the chatbot's contribution to their learning in six parameters, including, for example, whether the bot supported their learning process in the course, the extent to which it provided personalized responses to their questions and assisted them with assignments. Likert scale responses were scored on a scale of one to five. Cronbach's alpha reliability coefficients for the perceived contribution of

AI teaching assistants were 0.79. In addition, they answered two open-ended questions about their experience with the chatbot assistant: the advantages of using the course bot and the ways it supported their work, and the main challenges they encountered during use.

3.3. Procedure and data analysis

In all three groups, students participated in regular teacher education courses focused on classroom practice and lesson planning. The two experimental groups were given access to a locally developed artificial intelligence teaching assistant integrated into the course as a support tool for planning, writing, and refining lesson plans. This assistant was not available to the control group during the course. All participants completed the first two sections of the questionnaire regarding background information and perceived skill mastery, while participants in the experimental group also responded to the third section assessing the perceived contribution of AI teaching assistants to skill mastery.

As part of the data analysis procedure, descriptive statistics (means and standard deviations) were calculated for each dependent variable. Using chi-square tests, group differences in background characteristics (gender, age, teaching status, prior experience with artificial intelligence) were analyzed. A control variable was included in subsequent analyses where background variables were significant predictors of dependent measures.

Using one-way analyses of variance (ANOVA), as well as Scheffé post hoc tests, analyses were conducted to compare the overall use of artificial intelligence tools in the three groups (general-education control, general-education experimental, special-education experimental). The significance of group differences was then tested by using two-way ANOVAs after controlling for age. A further independent samples t-test was conducted to compare AI use between the general education groups (experimental vs. control) and the experimental groups (general vs. special education). As a final step, Pearson product-moment correlations were computed to determine how students reported receiving help from the AI assistant in relation to overall AI usage, as well as separately within each experimental group.

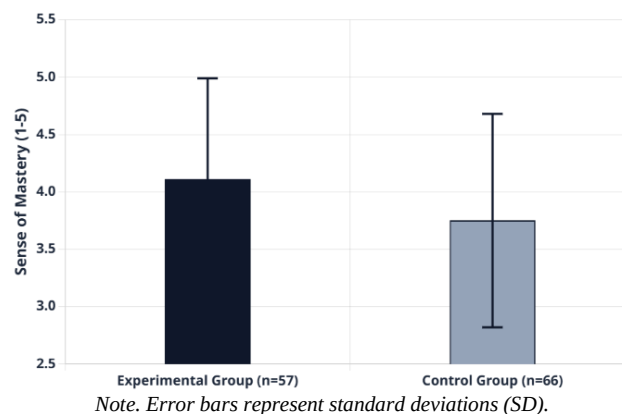
Open-ended responses from the two experimental groups were analyzed several times and coded inductively in order to identify recurring themes related to perceived benefits, challenges, and conditions for effective use of the chatbot assistant. These themes were then used to complement and interpret the quantitative findings regarding the use of artificial intelligence and the perceived contribution of the AI assistant.

4. Results

4.1. Quantitative findings

To examine the differences in the sense of mastery of skills between the experimental and control groups, independent samples T-tests were conducted. Significant differences were found between the groups in the skills of writing a complete lesson plan and writing operative objectives. These skills were higher among the experimental group that were taught using the AI chatbot assistant compared to the control group that did not take part in the study. Participants in the experimental group achieved significantly higher scores ($M = 4.11$, $SD = 0.88$, $n = 67$) compared to those in the control group ($M = 3.75$, $SD = 0.93$, $n = 57$), $t(121) = -2.15$, $p = .033$. Similarly, for writing operative objectives, the experimental group scored significantly higher ($M = 3.91$, $SD = 0.82$, $n = 66$) than the control group ($M = 3.55$, $SD = 1.00$, $n = 55$), $t(119) = 2.21$, $p = .029$. These results suggest that the use of AI teaching assistants had a positive impact on participants' performance, indicating that those who took part in the experimental group benefited more than participants who did not have the use of an AI assistant.

Figure 1. Sense of Mastery in Lesson Planning Skills: Experimental vs. Control Group.



Regarding the perceived contribution of the custom AI assistant, participants in both experimental settings reported high levels of satisfaction. Perceived utility scores were comparable between general education ($M = 3.97$, $SD = 0.79$) and special education ($M = 4.07$, $SD = 0.70$). This uniformity suggests that the assistant's design, which integrates a shared pedagogical corpus with course-specific materials, effectively supports diverse learning needs across different educational tracks.

Correlational analyses revealed distinct disciplinary patterns in how students engaged with the technology: In the general education experimental group, overall, AI use was positively correlated with the assistant's specific contribution to planning lesson flow ($r = .26$, $p < .05$). In the special education group, an additional significant positive correlation was identified between using AI to diversify learning styles and the students' critical monitoring of the assistant's outputs ($r = .44$, $p < .01$).

4.2. Qualitative insights

Thematic analysis identified three key dimensions of the student experience:

Prompt literacy as a professional competency: Participants emphasized that AI efficacy depended on instructional precision. Learning to "ask properly" specifying operational goals and target audiences reframed the AI as a draft-generator requiring active professional guidance rather than passive adoption.

Contextual alignment and personalization: Students valued the assistant's ability to "speak the language" of the course through tailored pedagogical ideas. However, they displayed critical awareness by identifying technical limitations, such as difficulties synchronizing general and operational goals or designing cohesive closing activities.

Shift from technical resource to epistemic partner: Perceptions evolved from viewing AI as a technical aid to a learning partner facilitating self-study. Aligned with the SRL framework, students maintained that the tool cannot replace human reasoning, advocating for a "human-in-the-loop" model where AI outputs remain subject to rigorous professional critique.

5. Conclusion

This study examined how a customized artificial intelligence teaching assistant may help teacher-education students develop lesson-planning skills. Quantitative findings indicated substantial promise in enhancing perceived mastery among students who used the assistant compared to a control group. Qualitative analyses revealed support for clarifying requirements, providing efficient feedback, refining prompting skills, and fostering critical evaluation of AI outputs. Results suggest that program-aligned artificial intelligence tools can improve self-regulated and personalized learning when used purposefully and in conjunction with explicit instruction in prompt literacy and critical AI literacy. Integrating such tools in coursework and clinical practice may enhance consistency between theory and application. Limitations include a single institutional context and reliance on self-report measures. Future research should examine long-term classroom impacts, disciplinary variations, and performance-based assessments. As a whole, AI tools can enhance professional preparation when used as reflective partners that complement, rather than replace, ethical and evidence-based pedagogical judgment.

References

- Aithal, P. S., & Aithal, S. (2024). Alternative applications of AI-based GPTs in teaching-learning and research-publications. *Tech Horizon 2023: Advancing Frontiers in Computer Science & Information Technology*, 77-87.
- Almasre, M. (2024). Development and evaluation of a custom GPT for the assessment of students' design in a typography course. *Education Sciences*, 14(148), 1-19. <https://doi.org/10.3390/educsci14020148>
- Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17), 12921. <https://doi.org/10.3390/su151712921>
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28, 973-1018.
- Lee, H. P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1-23).
- Masters, K., Benjamin, J., Agrawal, A., MacNeill, H., Pillow, M. T., & Mehta, N. (2024). Twelve tips on creating and using custom GPTs to enhance health professions education. *Medical Teacher*, 46(6), 752-756. <https://doi.org/10.1080/0142159X.2024.2305365>
- Zimmerman, B. J. (2000). Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13-39). San Diego, CA: Academic Press. <http://dx.doi.org/10.1016/B978-012109890-2/50031-7>