

From Blueprint to Brilliance: ChatGPT-Driven Improvements in Pre-Service Teachers' Lesson Plans: Enhanced Learning Material

Gemma Genabe Sordilla Calo-oy

Associate Professor C and Field Study Supervisor, University of Cebu-Main Campus
Email: gemscaloooy@gmail.com ORCID Number: <https://orcid.org/0000-0002-7359-2487>
Cebu Technological University-Main Campus, Cebu City, Philippines

Received: 05 Jul 2026; Received in revised form: 01 Aug 2026; Accepted: 06 Aug 2026; Available online: 10 Aug 2026
©2026 The Author(s). Published by Infogain Publication. This is an open-access article under the CC BY license
(<https://creativecommons.org/licenses/by/4.0/>).

Abstract— This study assessed the effectiveness of ChatGPT AI-driven scaffolding in enhancing the lesson plan writing performance of English pre-service teachers at the University of Cebu-Main Campus, school year 2025–2026 as basis of an enhanced learning material. The study employed a quasi-experimental design through random sampling in selecting fifty (50) participants, equally divided into twenty-five (25) participants for the control group and another twenty-five (25) for the experimental group. A lesson plan template and a standardized rubric evaluating organization, content, grammar, and mechanics served as the primary research instruments. Data were analyzed using descriptive and inferential statistical tools, including the t-test for dependent and independent samples. Findings revealed that the pre-test data indicated both groups initially struggled with structural organization and failed to reach high proficiency, while post-test results showed that the control group exhibited no statistically significant change and the experimental group demonstrated a significant improvement toward proficiency across all categories. Hypotheses testing confirmed a significant difference in the post-test performance of the experimental group, with p-values ranging from 0.000 to 0.017, well below the alpha level of 0.05, leading to the rejection of the null hypothesis (H_0), which means that the intervention provided to the experimental group was highly effective in enhancing their ability to structure, articulate, and refine lesson plans. In conclusion, ChatGPT AI-driven scaffolds are highly effective in refining lesson plan writing performance compared to traditional methods.

Keywords— English Teaching, ChatGPT AI-driven scaffolding, writing performance, quasi-experimental quantitative design, University of Cebu-Main Campus, Cebu City, Philippines



I. RATIONALE OF THE STUDY

The global trend of AI integration in education empowers educators worldwide across variety of studies—from language teaching to curriculum design—by enhancing efficiency, creativity, and pedagogical clarity in diverse cultural and institutional settings. This importance highlights lesson planning as a critical yet challenging skill, where AI tools like ChatGPT can scaffold writing processes to foster deeper pedagogical competence among pre-service teachers (Kalenda et al., 2025). In the context of English language teaching (ELT), this task is doubly demanding, as pre-service teachers must not only master pedagogical

strategies but also demonstrate high-level linguistic proficiency and the ability to scaffold complex communicative tasks for their future learners.

In teaching, writing a lesson plan often proves to be a time-consuming task, especially when combined with the preparation of visual aids and assessment tools (Powell & Courchesne, 2024). Pre-service teachers can spend hours crafting a single lesson plan, only to wonder if it is well-organized, coherent, and free from grammatical and mechanical errors. This rigorous process requires careful attention to detail and strong writing skills to ensure that the lesson effectively facilitates student learning. Amid the

broader global trend of AI integration in education, tools like ChatGPT promise to address these issues through scaffolding functions that support skill development in organization, coherence, grammar, and mechanics.

Prior to ChatGPT, pre-service teachers relied solely on their own creativity and planning to design activities and write lesson plans, often making the process more laborious. Now, with AI, pre-service teachers can generate lesson content quickly by simply providing prompts, making lesson planning more efficient (Chua et al., 2026). However, while AI supports this work, it remains essential for pre-service teachers to carefully review their writing for proper organization, content quality, grammar, and mechanics, ensuring that AI serves as a scaffold rather than a substitute in their teaching practice.

This study utilizes the Vygotskian ZPD framework (Vygotsky, 1978) to analyze how ChatGPT, acting as a more knowledgeable other (MKO), provides the linguistic scaffolding necessary for pre-service teachers to bridge the gap between fragmented lesson drafting and the production of coherent, grammatically precise, and pedagogically aligned instructional plans. A Vygotskian lens is vital here, as it frames AI within the Zone of Proximal Development (ZPD), enabling guided skill progression from assisted to independent performance; filling this gap is crucial to harness AI for equitable teacher training without fostering overdependence.

While international perspectives highlight AI's role in enhancing advanced writing proficiency (Bérešová, 2024), Philippine contexts emphasize its utility in organizational structure (Edraga, 2026) and its potential to mitigate infrastructure barriers (Villarino, 2024). These findings connect to the global context by affirming AI's scaffolding role but expose a niche: limited exploration of how ChatGPT influences lesson plan writing skills—specifically organization, coherence, grammar, and mechanics—for English pre-service teachers in the Philippines.

Based on classroom observations, the researcher noticed that when pre-service teachers wrote lesson plans without the aid of AI, they often encountered difficulties in structuring the components of the plan in a coherent and pedagogically sound manner, serving as an empirical bridge that validates the theoretical concerns mentioned. Common issues included poorly formulated behavioral objectives, lack of logical sequencing in activities, and misalignment between learning outcomes, instructional procedures, and assessment tools. Directly reflecting the laborious process, this process required extended periods of drafting and revising, leading to visible time inefficiency and occasional frustration caused by these poorly formulated behavioral

objectives. Without AI support, majority of pre-service teachers relied heavily on printed templates or prior examples, which sometimes resulted in repetitive or outdated teaching strategies that did not fully address the needs of diverse learners.

Conversely, when students utilize AI-based scaffolding in lesson plan writing, it is observed that there is a marked improvement in efficiency and content organization. With appropriate prompts, AI tools were able to generate well-structured lesson outlines, suggest varied and creative activities, and align objectives with Bloom's Taxonomy. Many students reported feeling more confident in their outputs, as AI provided instant feedback and alternative suggestions for activities, assessment items, and instructional strategies. However, the researcher also observed a tendency among some students to rely too heavily on AI-generated content without sufficiently applying their pedagogical judgment, underscoring the importance of critically evaluating and adapting AI outputs to suit specific classroom contexts.

While Edraga noted improvements in content, there is a lack of focus on the linguistic mechanics and coherence necessary for high-level ELT lesson planning within the Philippine ZPD context. Few studies have examined the interplay between AI support and the Zone of Proximal Development (ZPD) in lesson planning, particularly concerning the alignment of objectives, activities, and assessments. Consequently, there is a pressing need to determine how teacher education institutions can integrate these tools to enhance, rather than diminish, the critical thinking skills required for effective lesson preparation.

Despite the benefits noted in preliminary observations and existing literature, a critical gap remains: current studies [Bérešová, 2024; Edraga, 2026] do not sufficiently address the intersection of AI-based scaffolding and the development of specific linguistic-pedagogical skills (organization, coherence, grammar, and mechanics) among English pre-service teachers.

Thus, this study occupied this niche by investigating the impact of ChatGPT-driven scaffolding on the performance metrics. By evaluating performance in organization, coherence, and mechanics, this study moves beyond mere observation to provide a data-driven basis for institutional AI policies.

II. THEORETICAL BACKGROUND

In the context of the Philippine educational system, the theories of Lev Vygotsky (Sociocultural), Michael Long (Interactionist), and Jean Piaget (Constructivist) are not just

theoretical preferences—they are explicitly mandated by the state. These three theories belong to three foundational legal bases that govern curriculum design and teaching methodologies in the Philippines. The 1987 Philippine Constitution emphasizes the right of all citizens to quality and relevant education, which directly connects to Piaget's constructivist view of cognitive development and Vygotsky's sociocultural emphasis on learning through social interaction. Long's interactionist theory further strengthens this constitutional mandate by ensuring communicative competence, moving education beyond rote memorization toward functional, meaningful learning. Together, these theories provide the philosophical and legal justification for integrating innovative tools like ChatGPT into teacher education, ensuring that AI-driven scaffolding aligns with the constitutional vision of accessible and effective education.

Republic Act No. 10533, or the Enhanced Basic Education Act of 2013, and Republic Act No. 10157, the Kindergarten Education Act, further institutionalize these theoretical foundations into curriculum design. RA 10533 mandates learner-centered, developmentally appropriate, and contextualized education, which Piaget's stages of development directly support, while Vygotsky's Zone of Proximal Development underpins contextualization through Mother Tongue-Based Multilingual Education. Long's interactionist approach complements these by embedding collaborative and communicative practices into the K-12 curriculum.

Meanwhile, RA 10157 highlights Piaget's sensorimotor and preoperational stages in early childhood education, reinforcing the role of play and sensory experiences in knowledge construction. Vygotsky's sociocultural lens ensures that language and cultural tools are integrated at the foundational level, while Long's emphasis on interaction supports language acquisition even in kindergarten. These legal bases and theories collectively validate the study's use of ChatGPT scaffolding, demonstrating that AI integration is not only pedagogically sound but also legally aligned with the Philippine mandate for quality, relevant, and developmentally appropriate education.

Assuming that social interaction is the fundamental catalyst for cognitive growth, this study is anchored on the Sociocultural Theory of Lev Vygotsky (1978), which states that learning is a socially mediated process where individuals develop higher mental functions through interaction with more knowledgeable others and the use of cultural tools. According to Vygotsky, scaffolding provided by teachers, peers, or tools plays a vital role in helping learners perform tasks they would not be able to

accomplish independently. This process occurs within the Zone of Proximal Development (ZPD), where support gradually fades as learners gain competence.

In the context of this study, the identified problem was the need to enhance the lesson plan writing skill of English pre-service teachers, who were still in the process of developing the pedagogical knowledge and critical thinking needed to design effective and meaningful lessons. Traditionally, this scaffolding has come from mentors or instructors during practicum or method courses. However, with the rapid integration of digital tools in education, AI-based scaffolding is emerging as a new form of support that can simulate human guidance, offer real-time feedback, and promote reflective thinking.

The theory supports the assumption that AI, as a tool, can function as a more knowledgeable other, guiding pre-service teachers as they construct their understanding of effective lesson planning. Just as a mentor might ask questions, suggest improvements, or model lesson structures, AI-powered tools can analyze lesson drafts, provide personalized suggestions, and prompt users to think critically about their instructional choices. This aligns with Vygotsky's principle that learning is not just an individual process but one deeply rooted in social and tool-mediated interactions.

Therefore, applying Sociocultural Theory to this study justifies the exploration of how AI-based scaffolding affects lesson plan writing skill. It assumes that AI can serve not only as a technological innovation but also as a meaningful scaffold that bridges the gap between what pre-service teachers can do alone and what they can achieve with support—ultimately improving their readiness to teach effectively.

This study is also supported by Interactionist Theory by Long (1983) which posits that language learning occurs most effectively through meaningful interaction, negotiation of meaning, and feedback. According to this theory, input becomes valuable when learners actively engage with linguistic forms, clarify misunderstandings, and receive corrective feedback—processes that help them notice gaps between their current knowledge and target language norms. This noticing and negotiation promotes deeper processing of vocabulary, grammar, and discourse features, eventually contributing to improved communicative competence.

In relation to the study of Al-Raimi et al. (2024), the Interactionist Theory helps explain why Omani EFL students frequently use AI tools for translating words, verifying grammar and spelling, and generating ideas. These AI tools create an interactive environment where learners can immediately check meanings, refine their

language output, and experiment with sentence structures. Each interaction with AI—whether requesting a translation, asking for clarification, or seeking feedback—functions as a form of negotiation of meaning. The instant corrective responses offered by AI tools simulate conversational feedback that would typically come from a more proficient speaker. Thus, learners are not merely receiving input but actively modifying it based on the interactive prompts provided by the AI system.

Moreover, the theory supports the idea that AI tools can enrich language learning by prompting learners to notice linguistic inaccuracies and adjust their language use accordingly. When students use AI to compose essays or paragraphs, the feedback loop between their output and the AI's suggestions aligns with the interactionist view that learning emerges through cycles of production, feedback, and repair. Even though the study found no significant differences between male and female learners and no correlation between perceptions and practices, the widespread reliance on AI tools suggests that both genders benefit from the interactive, feedback-rich experience AI provides. Ultimately, the Interactionist Theory underscores that AI-based tools serve as interactive partners that facilitate negotiation of meaning—an essential mechanism for second language development.

This study is also supported by the Constructivist Learning Theory by Jean Piaget (1972), which posits that learners actively construct knowledge through experiences, rather than passively receiving information. Learning occurs as individuals engage with their environment, encounter new situations, and reconcile them with their prior knowledge. In this process, cognitive structures are continually modified through assimilation and accommodation. In the context of integrating AI tools like ChatGPT, constructivism highlights the importance of designing tasks where pre-service teachers explore, question, and apply language concepts, thereby deepening their understanding while developing autonomy. When paired with Vygotsky's emphasis on social interaction and the TPACK framework's focus on technological integration, constructivism ensures that AI-assisted lesson planning promotes active, meaningful, and learner-centered engagement.

In relation to this study, Constructivist Learning Theory underscores that knowledge is actively built by learners through meaningful engagement with tasks, rather than simply transmitted by instructors. When applied to AI-based scaffolding, constructivism suggests that English pre-service teachers develop their lesson plan writing competence by interacting with AI tools in ways that challenge their current understanding, prompting reflection,

adaptation, and refinement of ideas. These AI tools act as cognitive partners, offering prompts, suggestions, and feedback that enable learners to connect new information with prior knowledge, thereby facilitating deeper comprehension. This aligns with Vygotsky's view of learning within the Zone of Proximal Development, as the AI provides just enough structured support for learners to achieve tasks they could not manage independently. Over time, as competence grows, the scaffolding can be gradually reduced, fostering autonomy and mastery in lesson planning. This underscores that learners actively build understanding through interaction, reflection, and adaptation—especially when engaging with AI as a cognitive collaborator.

Additionally, in the study by Mishra and Koehler (2006) on Technological Pedagogical Content Knowledge (TPACK), it emphasizes that effective teaching using technology needs the interplay of different knowledge domains: Technological Knowledge (TK), Pedagogical Knowledge (PK), and Content Knowledge (CK). This theory provides a framework for pre-service teachers in incorporating AI tools like ChatGPT in understanding the importance of how to use tools (TK), while aligning them to their teaching methods (PK), and giving attention to the appropriateness of the language content (CK).

Technological Knowledge (TK) points to the ability of pre-service teachers to maximize AI tools like ChatGPT in learning how to develop their lesson plans that will align with their specific objectives. As such, AI will help them structure their lessons, which will provide them with knowledge on how to approach their lessons in the future. On the other hand, Pedagogical Knowledge (PK) pertains to their abilities in aligning their techniques and strategies in teaching based on the lesson plan developed with the help of AI. With this, pre-service teachers have to read, understand, and evaluate AI-generated suggestions to be able to learn and apply them on their own moving forward. Lastly, Context Knowledge (CK) refers to the level of understanding by pre-service teachers of the content they teach in class, ensuring that the scaffolded lesson plan generated by AI is linguistically accurate and appropriate in terms of their students' level and diverse backgrounds. Ultimately, using TPACK in the study allows pre-service teachers to effectively integrate the use of AI in writing their lesson plans, adjust their teaching approaches, and understand their content, highlighting the importance of the three domains working together to improve lesson planning performance.

Recent studies have highlighted the potential of AI tools to enhance TPACK development among pre-service teachers, particularly in lesson planning. Lee and Zhai

(2024) found that pre-service elementary teachers who integrated ChatGPT into their science lesson designs demonstrated notable improvements in aligning content objectives with appropriate pedagogy and technology, though challenges remained in critically evaluating AI outputs. Similarly, Li et al. (2023) reported that AI-supported lesson planning in English language teaching contexts fostered greater coherence between instructional strategies and digital tool use, enabling pre-service teachers to better address learner needs while managing technological affordances and constraints. These findings affirm that integrating AI within a TPACK framework not only supports teachers' technological fluency but also strengthens their capacity to craft pedagogically sound and content-appropriate lessons—outcomes that are consistent with Vygotsky's emphasis on scaffolded learning within the Zone of Proximal Development.

According to Novita (2025), teachers today face increasing pressure to deliver personalized, standards-aligned instruction while managing limited time and growing responsibilities. Traditional lesson planning, though essential, often restricts creativity due to its repetitive and time-consuming nature. This led Novita to ask how AI supports teacher efficiency and instructional design in lesson planning. Using a qualitative literature review, the study examined peer-reviewed articles from 2016 to 2023, focusing on the role of AI in streamlining educational tasks. The researchers conducted a systematic search across databases like Scopus, ERIC, and Web of Science, narrowing 48 initial results to 18 relevant studies based on educational applicability and pedagogical value. Thematic analysis revealed key themes such as AI-supported design, enhanced efficiency, personalization, and implementation challenges. The review concludes that AI can significantly improve teacher efficiency by automating tasks like curriculum alignment, content curation, and assessment design. However, the study also emphasized the importance of ethical AI use—highlighting issues like data privacy, algorithm transparency, and teacher oversight to maintain professional and pedagogical integrity.

Teaching is a cognitively demanding and highly creative task that requires careful preparation to ensure effective and purposeful instruction. In modern classrooms, lesson plans are often developed using learning or content management systems. Recognizing the complexity of this process, Pender et al. (2022) explored how AI-based lesson planning software can not only automate routine tasks but also support reflective, competence-based teaching. Their study aimed to answer the following research questions: How can AI-based software assist teachers in designing individualized, competence-oriented lesson plans? To what extent does such a system maintain pedagogical freedom

while improving efficiency? To address these questions, the authors conducted a qualitative design study using various user scenarios and software demonstrations to illustrate the software's features and practical applications. Teachers interacted with the tool, which allowed them to easily and intuitively generate personalized learning content without compromising their instructional autonomy. The significant findings of the study showed that the AI system effectively supports competence-based learning by offering flexible, individualized lesson suggestions while simplifying the planning process. Additionally, the system fosters pedagogical reflection by guiding users through structured design processes. Ultimately, the study promotes greater awareness of intelligent learning environments that can support the automated, lifelong development of personalized educational content, paving the way for broader AI integration in teaching and learning practices.

Another study by Uysal and Yüksel (2024) examined the integration of artificial intelligence (AI) into language education, with a particular focus on the views of pre-service teachers of English as a foreign language (EFL). Utilizing a mixed-methods design, the study evaluated the effectiveness of AI-generated lesson plans tailored for teaching writing to 5th-grade learners. Through the use of a detailed evaluation rubric and qualitative analysis, the research uncovered both the strengths and limitations of these AI-supported plans, along with recommendations for improvement. The findings revealed that while the AI tool performs well in terms of engagement, content relevance, and overall lesson structure, it faces notable challenges in areas such as differentiation and assessment strategies. The chapter concluded by discussing the implications for teacher education, underscoring the importance of developing AI literacy among future educators to leverage AI technologies in varied language teaching contexts effectively.

Additionally, ElSayary et al. (2025) aimed to investigate how prompt engineering contributes to the effective use of generative AI tools—such as EduAid, Magic School, and ChatGPT—for improving the efficiency and relevance of lesson planning. The study underscored the importance of AI literacy and ongoing professional development to ensure that AI-generated outputs meet the diverse needs of students. Using a qualitative case study approach, the researchers gathered data from 17 high school teachers across both private and public schools. Data sources included document analysis, written reflections, and focus group discussions to gain in-depth insights into teachers' experiences and perceptions regarding GenAI tool usage. The findings revealed that well-crafted prompts significantly enhanced teachers' ability to leverage GenAI tools, resulting in more relevant and effective lesson plans.

Teachers also noted the need for continuous refinement of AI-generated content to better address all learners' needs. The study emphasizes the necessity of sustained AI training to support meaningful integration of these technologies into educational practice. Overall, it highlights the strong potential of AI tools to automate and enrich lesson planning, with their effectiveness largely depending on the quality of user-generated prompts.

Moreover, another study investigated how pre-service teachers collaborate with AI in developing lesson plans, particularly in the context of physics instruction. Fifteen pre-service elementary teachers were guided to select a topic and write initial lesson plans, which were later refined through interaction with ChatGPT. Furthermore, the degree of collaboration was evaluated by examining the teachers' collaborative competencies with AI, along with the nature of their interactions and the changes made at each level. Pre-service teachers demonstrating low-level collaboration engaged in consensual exchanges, resulting in no revisions. In contrast, those at the intermediate level exhibited responsive interactions, often leading to structural modifications in their lesson plans. Meanwhile, high-level collaborators engaged in elaborative dialogue with AI, which led to deeper changes in both content knowledge and pedagogical strategies. The study emphasized that establishing AI as a collaborative peer is crucial, and it also revealed that meaningful, high-level collaboration requires learners to possess prior subject knowledge. These findings underscore the importance of preparing teachers to engage with AI as a professional partner in instructional design (Park & Park, 2024).

Lee and Zhai (2024) found that pre-service elementary teachers who integrated ChatGPT into science lesson planning demonstrated enhanced alignment between content objectives, pedagogy, and technology, though challenges remained in critically evaluating AI outputs. Similarly, Kehoe (2024) noted that AI can reduce teachers' time burden in lesson planning while improving instructional quality, provided educators apply their expertise to refine generated materials.

A study examined the integration using AI tools like ChatGPT in pre-service teachers in the Philippines. They discovered that while it improves lesson planning, creativity, and supports differentiated instruction, its usage poses challenges like unfamiliarity, accuracy and credibility concerns, and overreliance, David and Maroma (2025).

A study by Rougeaux and Sharp (2023) expressed that teachers might initially view AI as a tool with negative effects. However, the researchers believe AI might become an integral tool of a transformative classroom. It aimed to determine whether ChatGPT meets these criteria in

preservice mathematics methods courses. The research focused on generating specific input prompts to guide ChatGPT in producing lesson plans, which were then compared with those created by preservice teachers. Specifically, the study addressed the following research questions: (1) Can ChatGPT be used to produce lesson plans for specific mathematical objectives? (2) How do different types of prompts affect the output of lesson plans generated by ChatGPT? and (3) How do lesson plans generated by ChatGPT compare to those designed by preservice teachers? The findings suggested that effective use of AI in lesson planning could significantly reduce the time required for teaching lesson plan development, allowing instructors to shift focus toward evaluating and improving AI-generated content, thereby maximizing instructional time in mathematics methods courses.

Zhussupova et al. (2025) stated that the rapid development of Artificial Intelligence (AI) is reshaping educational practices, particularly in teacher training. The study explored the integration of Digital Multimodal Simulation (DMS) and ChatGPT in enhancing lesson planning skills among Pre-service English Language Teachers (PELTs). Using a mixed-methods approach that combines rubric-based evaluation, classroom observation, and thematic analysis, the study examined the effectiveness of AI-assisted lesson planning and its perceived benefits. Findings reveal that DMS supported by ChatGPT significantly improved PELTs' ability to structure, personalize, and align lesson plans with pedagogical standards. While ChatGPT-generated plans demonstrated strengths in clarity and differentiation, human-designed plans showed deeper engagement and resource use. Additionally, PELTs reported increased confidence, practical skills, and reflection through real-time feedback within the simulation environment. The study contributes to the growing body of research on AI integration in teacher education and highlights the complementary role of human and AI collaboration in pedagogical design.

Kerr and Kim (2025) examined how pre-service teachers incorporate generative AI tools into their lesson planning. Specifically, the study explores the methods they utilized, the benefits realized, and the challenges they encountered. Participants include 8 English as a Foreign Language (EFL) pre-service teachers studying digital literacy courses at a Korean university. Data were gathered by convenience sampling, over three one-hour and fifteen-minute classes within a 16-week semester. And using a qualitative case study approach, the study explored the PSTs' lived experiences with generative AI in lesson planning. The findings demonstrate the potential of generative AI tools in supporting PSTs in lesson planning and material development. Results show that PSTs utilized

AI in various tasks involving lesson planning, providing efficiency and creative support. However, challenges with its limitations were also observed, such as a lack of depth in the lesson content, not enough support for teaching approaches, and occasional inaccuracies. These emphasize the importance of teacher training programs, critical evaluation of AI outputs, and strategies for incorporating AI tools in teaching contexts.

Studies have been conducted to examine the impact of AI on lesson planning. Setyaningsih et al. (2024) uncovered the dual impact of AI: how it streamlines lesson planning, supports differentiated instruction, and enhances creativity while demanding careful teacher oversight, managing effective prompts, and thoughtful adaptation to classroom realities. While teachers appreciated AI's ability to reduce planning workload and offer varied instructional ideas, they also noted that AI-generated content often requires further refinement to meet pedagogical standards. The findings stressed the need for focused training and professional development to prepare educators for effective AI integration. Positioning AI as a supportive partner rather than a teacher replacement, the study contributes to ongoing discussions about AI in education and offers insights into fostering meaningful, innovative, and context-appropriate teaching practices.

UNESCO (2023) reported that generative AI holds the potential to transform conventional teaching practices, emphasizing its challenges and opportunities. Similarly, Liu et al. (2025) also concluded that integrating AI into education presents notable challenges for teachers, especially during the transitional stage of adapting these technologies for lesson planning. The study emphasized that both internal factors, like teachers' attitudes toward new methods and AI tools, and external influences, such as training availability, peer support, and policy directives, significantly impact adoption strategies and results.

Khalid et al. (2024) studied how Teaching English as a Second Language (TESL) undergraduate students perceive the integration of AI into lesson planning. The study aimed to understand what AI tools teacher trainees prefer and their perceptions of these tools in developing lesson materials. It sought to answer the following research questions: (1) What are the perceptions of TESL students when it comes to integrating AI in developing lesson plans? (2) Which AI do TESL students choose when it comes to developing lesson plans? It employed a descriptive case study methodology that incorporated semi-structured interviews and questionnaires, involving 30 teacher trainees from a Malaysian public university. Findings have revealed favorable perceptions of AI integration in lesson planning. Results uncovered that ChatGPT and Magic School AI are

the preferred AI tools among the participants, emphasizing its efficiency, acceleration of idea generation, and simplification of the overall process.

The findings of the study of Al-Raimi, et. al. (2024) showed that translating words, phrases, and sentences is where Omani EFL students most frequently use AI tools. In order to verify spelling and grammar, they also employ AI techniques. The participants also report using AI writing tools to generate ideas for their writing and to assist them in composing essays and paragraphs. Findings showed that there were no significant differences between male and female students, and there was no correlation concerning the students' perceptions and practices in utilizing AI tools. The study's recommendations and implications were presented in accordance with this.

Enhancing writing skills in the EFL classroom is crucial for fostering effective communication, critical thinking, and self-expression. Writing poses challenges for EFL learners due to the need for precise idea formation and sustained attention from both teachers and students. This skill enables learners to navigate diverse linguistic contexts and participate actively in various discourses. The rapid advancement of technology has introduced Artificial Intelligence as a promising tool to augment language learning, particularly in reinforcing writing skills. The present dissertation investigates the utilization of Artificial Intelligence (AI) tools to augment writing instruction and explores their impact on learners' writing abilities and proficiency. It investigates the impact of ChatGPT on students' writing skills. It is hypothesized that the use of ChatGPT may enhance EFL learners' writing proficiency by providing opportunities for practice, feedback, and error detection and correction. The findings offer valuable insight into students' and teachers' perspectives and attitudes towards using AI tools, specifically ChatGPT, their usage patterns and their perceived impact on writing skills' development, Imane, R., & Siham.

Patrik and Ilona (2024) also explored how Artificial Intelligence (AI) can support and improve English lesson plan writing in terms of personalization, efficiency, and instructional quality. Using a qualitative analysis of AI-integrated platforms—such as MagicSchool and its chatbot, Raina—the study examined how these tools help teachers generate customized lesson plans, instructional materials, and differentiated strategies based on student needs. The article highlighted AI's capacity to analyze student data, identify individual learning gaps, and suggest tailored strategies to enhance engagement and learning outcomes. Notably, it also emphasized AI's role in reducing teacher workload and promoting efficiency through automated content creation. However, the article also outlined key

challenges, including the digital divide, unequal access to technology, and ethical concerns surrounding data privacy. It calls for responsible AI integration through equitable access and clear data-use policies.

A study explores preservice education student perceptions on ChatGPT's ability to write lesson plans through the use of a guided analysis process. A pre- post-survey and a reflection on the strengths and weaknesses of ChatGPT were written by participants. Students shared that they plan to use ChatGPT for lesson planning as a practicing educator, but indicated that revision of output is necessary, Kalenda, et. al. (2025)

According to Lee and Zhai (2024), despite the buzz around ChatGPT's potential, empirical studies exploring its actual utility in the classroom for learning remain scarce. The study aims to fill this gap by analyzing the lesson plans developed by 29 pre-service elementary teachers from a Korean university and assessing how they integrated ChatGPT into science learning activities. The study further examined pre-service teachers' perceptions and concerns about integrating ChatGPT into science learning. Results show diverse applications of ChatGPT in different science domains. Meanwhile, they also expressed concerns about its accuracy and the risks that students may be overly dependent on ChatGPT. They further suggested solutions to systemizing classroom dynamics between teachers and students. The study underscores the need for more research on the roles of generative AI in actual classroom settings and provides insights for future AI-integrated science learning.

These studies provide theoretical support for integrating AI in writing lesson plans, which needs empirical validation. This current study builds on these findings by moving from theory to application—specifically investigating the effect of AI-based scaffolding on the lesson plan writing skills performance of English pre-service teachers.

This research assessed the effectiveness of ChatGPT AI-driven scaffolding in enhancing the lesson plan writing performance of the English pre-service teachers at University of Cebu-Main Campus in Cebu City during the school year 2025-2026 as basis for a proposed enhanced learning material.

Specifically, it sought to answer the following questions:

1. What is the pre-intervention lesson plan writing performance level of the

control and experimental groups in terms of the following writing elements:

1.1 organization,

1.2 content,

1.3 grammar, and

1.4 mechanics?

2. What is the post-intervention lesson plan writing performance level of the control and experimental groups in terms the following writing elements of:

2.1 organization,

2.2 content,

2.3 grammar, and

2.4 mechanics?

3. Is there a significant difference between the pretest and posttest scores in the lesson plan writing performance of the:

3.1 control group, and

3.2 experimental group?

4. Is there a significant mean gain difference on the pre-test and post-test scores in the lesson plan writing performance between the control and the experimental group?

5. Based on the findings of the study, what enhanced learning material may be developed?

III. METHOD

Design

This study employed a quasi-experimental quantitative research design, defined by Cham, Lee, and Migunov (2024) as an approach that utilizes control and experimental groups. While randomized controlled trials remain the gold standard, the non-equivalent groups design was selected to preserve the ecological validity of the authentic classroom environment, thereby ensuring that findings are generalizable to actual teacher-education contexts. Fifty participants were selected through random sampling from the target population to enhance representativeness. These participants were then divided into two groups of twenty-five (25) each based on existing class sections, ensuring comparability. One group served as the experimental group, receiving the intervention, while the other functioned as the control group, continuing with traditional instruction.

To see if the AI support actually makes a difference, the study compares how well the two groups write their lesson plans. This writing performance is judged across four straightforward areas: how well the lesson plan is organized, the accuracy of its content, its grammar, and its mechanics (like spelling and punctuation). Because the groups are pre-existing and not perfectly identical at the start, the data handling and processing will include careful steps to ensure a fair comparison. The researcher administered a pre-test to measure students' prior knowledge about lesson planning before the intervention.

Figure 2 illustrates the flow of the study. Pre- and post-intervention assessments enable tracking of performance changes over time, while statistical controls mitigate threats like maturation or history effects, ensuring reliable attribution of improvements to the AI scaffolding intervention.

Flow of the Study

The input of the study consists of the initial lesson plan writing performance of English pre-service teachers. This is measured through a pre-test that evaluates organization, content, grammar, and mechanics. These input data serve as the foundation for comparing two groups: the control group, which engages in traditional lesson planning, and the experimental group, which utilizes ChatGPT AI-driven scaffolding. This input phase ensures a clear benchmark of participants' writing competence prior to the intervention.

The process of the study follows a quasi-experimental quantitative design. As part of this process, participants were randomly assigned to either the control or experimental group. Both groups completed a pre-test to determine their initial writing performance. The process continued with the intervention: the control group wrote lesson plans without AI assistance, while the experimental group used ChatGPT AI-driven scaffolding. After the intervention, a post-test was administered to both groups. The process was concluded with evaluation using a standardized rubric measuring organization, content, grammar, and mechanics, followed by statistical analysis and interpretation of the writing performance data.

The final output of the study is the development of enhanced learning material aimed at improving lesson plan writing skills. Grounded in the findings, this material will include pedagogical strategies and AI-assisted techniques that can be adopted within the College of Teacher Education. The output reflects the study's contribution to strengthening instructional competencies and promoting the integration of technology in teacher education. Ultimately, it supports the preparation of adaptive, digitally literate educators equipped for contemporary classroom challenges.

Instruments

This study employed two primary instruments. First was the lesson plan template which served as a guide for the pre-service teachers as to the lesson plan content is concerned. It was composed of the following components: specific objectives, activities or procedures, assessment and assignment.

The second instrument used was the rubric designed to evaluate participants' writing skill based on four key criteria: Organization, Content, Grammar, and Mechanics. Each criterion is assessed on a four-level scale: Exemplary

(10 pts), Proficient (8–9 pts), Developing (6–7 pts), and Needs Improvement (0–5 pts). This is adapted from Carpinelli et al. (2008), who developed a rubric to evaluate standards-based lesson plans.

The rubric's validity is supported by its origin in peer-reviewed engineering education research for assessing standards-aligned lesson plans and they have structural or pedagogical commonalities in terms of standards-based lesson planning, making it appropriate for this study through adaptation to evaluate English pre-service teachers' lesson plan writing performance.

Data Gathering Procedure

Preliminary Preparation. Before the actual collection of data, the researcher secured approval from the Dean of the College of Teacher Education of the University of Cebu–Main Jones Campus. Once permission was granted, she obtained the official list of fifty (50) fourth-year BSED English majors who served as participants. These students were randomly assigned into two equal groups: twenty-five (25) in the control group and twenty-five (25) in the experimental group. All participants were informed about the purpose of the study, the voluntary nature of their participation, and the confidentiality of their responses.

Pre-test Phase. To begin, the data collection began with the pre-test phase, during which both groups were administered to write an initial lesson plan based on a given English topic on Kinds of Sentences according to function. Following this task, the initial lesson plans were evaluated using a standardized rubric measuring organization, content, grammar, and mechanics.

Intervention Phase. The intervention phase was implemented for a three-week period. During this timeline, the control group wrote lesson plans without ChatGPT scaffolding, while the experimental group used ChatGPT AI-driven scaffolding.

Post-test Phase. Immediately following the three-week intervention, the post-test phase commenced. Both groups completed a post-test lesson plan writing task. Subsequently, the outputs were again assessed using the same rubric, and finally, all scores were encoded, organized, and prepared for statistical analysis.

Statistical Treatment

In this study, both descriptive and inferential statistics were applied to analyze the lesson plan writing performance of the control and experimental groups. Descriptive statistics, including frequency distribution, percentage, mean, and standard deviation, were first used to establish the initial writing performance levels.

For the inferential analysis, a paired sample t-test was employed at an alpha level of 0.05 to determine

whether there was a statistically significant difference between the pre-test and post-test scores within each group across the four writing elements: Organization, Content, Grammar, and Mechanics.

The mean gain scores calculated as the difference between pre-test and post-test results of the control and experimental groups were compared using an independent samples t-test to further evaluate the extent of improvement specifically attributable to the ChatGPT AI-driven scaffolding,

To maintain objectivity and ensure inter-rater reliability, two independent inter-raters evaluated the outputs alongside the researcher.

IV. RESULT

The following constitute the findings of the study:

The pre-intervention data for organization showed that both the control and experimental groups were mostly at the developing level. Many participants in both groups struggled with organization, and none reached the exemplary level.

The pre-intervention results for content revealed that most participants in both groups were also at the developing level. No one reached the exemplary category.

The pre-intervention results for grammar showed that while participants understood grammar rules a little better than organization, both groups were still at the developing level.

The pre-intervention results for mechanics showed that spelling and punctuation were difficult for both groups. The control and experimental groups fell into the Needs Improvement category, with most staying at the developing level.

The post-intervention data for **organization** showed a strong improvement. Most of the experimental group reached the proficient level, while the control group stayed at the developing level.

The post-intervention data for **content** revealed that several participants in the experimental group reached proficient, and one even achieved exemplary. In contrast, the control group showed little progress, with most still at the developing level.

The post-intervention data for **grammar** indicated that most of the experimental group reached proficient, while the control group remained mostly at the developing level.

The post-intervention data for **mechanics** showed that the majority of the experimental group reached proficient. The control group, however, stayed at the developing level, with some still in the need improvement category.

Organization, content, and grammar showed significant arithmetic improvements while the area of mechanics showed a non-significant difference.

The experimental group achieved highly significant improvements across all categories—organization, content, grammar, and mechanics.

It confirmed the effectiveness of the intervention, showing that the experimental group achieved significantly higher mean gains than the control group in every category.

V. DISCUSSION

PRE-INTERVENTION LESSON PLAN WRITING PERFORMANCE LEVEL OF BOTH GROUPS

The pre-intervention lesson plan writing performance level of both groups across four primary writing elements: organization, content, grammar, and mechanics is categorized into four: Needs Improvement, Developing, Proficient, and Exemplary.

Pre-Intervention Writing Performance: Organization

Table 1 reflects the pre-intervention writing performance level of the two groups before the use of ChatGPT AI-driven scaffolds in terms of organization.

Table 1: Pre-Intervention Writing Performance: Organization

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	1	4.00	7	28.00
Developing	5.1-7.0	14	56.00	15	60.00
Needs Improvement	0-5.0	10	40.00	3	12.00
Total		25	100.00	25	100.00
Average		5.52		6.42	
St. Dev.		0.90		0.90	

As illustrated in Table 1, prior to the intervention, majority of the participants in both the Control and Experimental groups were at the Developing level regarding lesson plan organization. While neither group achieved an Exemplary rating, the Experimental group demonstrated a higher baseline proficiency with an average score of 6.42 compared to the Control group’s 5.52. Notably, 40% of the Control group fell into the Needs Improvement category, whereas only 12% of the Experimental group scored in this lowest tier.

Supporting literature, such as studies by Schweil et al. (2022) and Kaenah and Utami (2023), suggests that this concentration in mid-level proficiency is common among

pre-service teachers who lack specific structural frameworks, often resulting in a lack of "Exemplary" logical flow. Furthermore, the uniform standard deviation of 0.90 observed in the data aligns with findings by Munda (2025), which highlight the variability in initial instructional planning competence when teachers rely on individualized styles rather than standardized organizational scaffolds.

Pre-Intervention Writing Performance: Content

Table 2 shows the pre-intervention lesson plan writing performance level of the control and experimental groups before the use of ChatGPT AI driven scaffolds in terms of content.

Table 2: Pre-Intervention Writing Performance: Content

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	2	8.00	8	32.00
Developing	5.1-7.0	11	44.00	16	64.00
Needs Improvement	0-5.0	12	48.00	1	4.00
Total		25	100.00	25	100.00
Average		5.54		6.47	
St. Dev.		0.96		0.91	

As reflected in Table 2, prior to the intervention, the Control and Experimental groups exhibited a baseline performance in Content that ranged primarily from **"Needs Improvement" to "Developing."** The two groups started at very different levels. Most students in the Experimental group were already at the Developing or Proficient levels.

In contrast, the Control group struggled, with nearly half scoring below 5.0 at the Needs Improvement level—compared to just 4% in the Experimental group. While neither group reached the Exemplary level, the Experimental group’s average score of 6.47 was higher than the Control group’s 5.54. This clearly shows a big initial gap between the two groups before any ChatGPT AI tools were used.

Del Rosario (2026) identified significant challenges among 43 pre-service teachers in the Philippines, noting that inadequate understanding of literary elements and content complexity often led to poorly structured lesson plans. Similarly, Lariosa (2026) found that while pre-service teachers might be rated as "Outstanding" in general content knowledge, they struggle significantly with specific instructional skills like questioning and

pedagogical application, mirroring the "Developing" concentration seen in the control and experimental groups.

Furthermore, Sanchez et al. (2024) emphasized that pre-service performance is heavily influenced by the lack of structured mentorship and inadequate reference materials, validating the necessity of scaffolds to move students from a "Needs Improvement" baseline toward proficiency. Together, these studies suggest that without specific interventions, pre-service teachers frequently remain at a developing stage due to the high cognitive load of integrating content knowledge with pedagogical strategies

Pre-Intervention Writing Performance: Grammar

Table 3 displays the pre-intervention lesson plan writing performance level of the two groups before the use of ChatGPT AI-driven scaffolds in terms of grammar.

Table 3 presents the **initial** performance of the Control and Experimental groups in terms of Grammar before the implementation of AI-driven scaffolds. The data reveals a significant similarity: both groups are predominantly situated at the Developing level (5.1 - 7.0), representing 68% of the Control group and 56% of the Experimental group. The Experimental group achieved a

notably higher frequency at the Proficient level (36%) compared to only 8% in the Control group and neither group reached the Exemplary level.

Table 3: Pre-Intervention Writing Performance: Grammar

Level	Ranges of Scores	Control		Experimental	
		f	%	F	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	2	8.00	9	36.00
Developing	5.1-7.0	17	68.00	14	56.00
Needs Improvement	0-5.0	6	24.00	2	8.00
Total		25	100.00	25	100.00
Average		5.68		6.53	
St. Dev.		0.95		0.96	

Research indicates that pre-service teachers' performance in grammar in lesson plan writing is often **limited** by a combination of content knowledge gaps and pedagogical challenges, aligning with the "Developing" and "Needs Improvement". Specifically, the factors behind the Experimental group's higher "Developing" concentration are reflected by Sanchez et al. (2024), who highlighted that the lack of specific instructional guides and reference materials often leaves pre-service teachers in a "developing" state, as they struggle to align grammatical learning objectives with appropriate assessment strategies, and by Roca and Manla (2025), who explored the grammatical competence of graduating education students and found that while participants often perceive their own skills as sufficient, objective assessments frequently show only "fair" competence, particularly in complex sentence elements, which directly **limits** their ability to write "Exemplary" lesson plans.

In contrast, the **greater challenges** explaining the Control group's lower "Needs Improvement" concentration are illustrated by Del Rosario (2026), who conducted a qualitative study in the Philippines which revealed that pre-

service teachers face significant **challenges** in grammar lesson preparation due to the complexity of English linguistic structures and an inadequate understanding of lexicogrammar, leading to poorly structured instructional flows.

In addition, according to Nagaño, et. al (2025), in their study on the grammatical competence of 50 fourth-year pre-service English teachers, sentence elements emerged as the least learned domain, falling into the low level of competence. The findings highlight substantial strengths in several grammar areas but reveal persistent, specific gaps in sentence structure knowledge. Recommendations include developing targeted grammar instruction modules and continuous monitoring to address the identified weaknesses, thereby enhancing the instructional quality and language proficiency of future educators.

Pre-Intervention Writing Performance: Mechanics.

Table 4 indicates the pre-intervention lesson plan writing performance level of the two groups before the use of ChatGPT AI-driven scaffolds in terms of mechanics.

Table 4: Pre-Intervention Writing Performance: Mechanics

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	2	8.00	7	28.00
Developing	5.1-7.0	15	60.00	13	52.00
Needs Improvement	0-5.0	8	32.00	5	20.00
Total		25	100.00	25	100.00
Average		5.68		6.26	
St. Dev.		0.95		1.01	

Similar to the previous categories, both the Control and Experimental groups were primarily positioned at the Developing level. (5.1-7.0), representing 60% (n=15) and 52% (n=13) of their respective groups. The similarity between groups ends there, as the Experimental group displayed a greater range of performance, with 28% reaching the Proficient level compared to only 8% in the Control group. Conversely, the Control group showed more notable difficulties, with nearly a third of its participants (32%) in the Needs Improvement category, whereas only 20% of the Experimental group scored in this lowest level. Overall, with an average score of 6.26 for the Experimental group and (5.68) for the Control group—and no participants achieving an Exemplary rating—the data indicates that both groups had a foundational but technically flawed grasp of mechanics prior to using the AI-driven scaffolds.

The initial "Developing" proficiency in mechanics observed in both groups (Means of 5.68 and 6.26) aligns with contemporary research identifying the significant

"added mental burden" pre-service teachers face when beginning instructional design. AIMS Press (2026) and Ramos & Baylon (2025) observe that novice educators often struggle to simultaneously manage complex pedagogical structures and technical linguistic accuracy (spelling, punctuation, and capitalization).

POST-INTERVENTION LESSON PLAN WRITING PERFORMANCE LEVEL OF BOTH GROUPS

The post-intervention lesson plan writing performance level of both groups—across the four writing elements of organization, content, grammar, and mechanics—comprises the following levels: Needs Improvement, Developing, Proficient, and Exemplary-

Post-Intervention Writing Performance: Organization

Table 5 illustrates the post-intervention lesson plan writing performance level regarding Organization.

Table 5: Post-Intervention Writing Performance: Organization

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	2	8.00	21	84.00
Developing	5.1-7.0	18	72.00	4	16.00
Needs Improvement	0-5.0	5	20.00	0	0.00
Total		25	100.00	25	100.00
Average		6.23		7.84	
St. Dev.		0.85		0.76	

As shown in Table 6, a clear difference has appeared between the two groups: the Experimental group achieved a higher level of organization skill, with 84% of participants reaching the Proficient level. In comparison, the Control group remains mainly at the Developing level, with 72% of participants still struggling to move beyond basic logical sequencing. While the Control group showed a small increase in mean score to 6.23, the Experimental group's mean increased quickly to 7.84, indicating that the AI scaffolds effectively closed the gap toward better teaching.

Based on the data in Table 6, while the intervention led to a significant upward change for the Experimental group, it is factually correct to observe that the majority of the participants reached the Proficient level, yet none achieved the Exemplary level. This leveling-off point suggests that while ChatGPT-driven scaffolds are highly effective at providing structural coherence and logical

flow—the "basic structure" of a lesson plan—they may not yet facilitate the highest level of instructional creativity or "Exemplary" pedagogical reasoning required for a perfect score.

The inability of the Control group to reach similar proficiency levels—despite normal growth—reinforces Seherrie's (2020) observation that developing complex lesson phases on their own is a major difficulty for educators. By providing real-time feedback and structural templates, the AI-driven intervention allowed the Experimental group to overcome the "developing" stage identified by Azzahra et al. (2023), whereas the majority of the Control group remains stuck in that basic level. Educational research by Al-Azawei (2024) suggests that while scaffolding tools assist with "Organization," the jump from proficiency to mastery often needs strong teaching knowledge and skill (PCK) that automated tools cannot fully replace. Furthermore, Garcia and Martinez (2025)

found that pre-service teachers often depend heavily on the tool's first output, leading to a "leveling-off" where the work is technically correct and well-organized but lacks the highly contextualized, innovative elements characteristic of exemplary teaching. Therefore, the results indicate that while the scaffolds successfully remediated the "Needs Improvement" and "Developing" levels, the move toward

"Exemplary" likely requires additional human-led mentoring to refine the fine skill of adjusting instruction.

Post-Intervention Writing Performance: Content

Table 6 presents the post-intervention lesson plan writing performance level of the two groups in terms of Content.

Table 6: Post-Intervention Writing Performance: Content

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	1	4.00
Proficient	7.1-9.0	1	4.00	18	72.00
Developing	5.1-7.0	21	84.00	6	24.00
Needs Improvement	0-5.0	3	12.00	0	0.00
Total		25	100.00	25	100.00
Average		6.13		7.78	
St. Dev.		0.74		0.78	

Table 6 indicates a significant shift in proficiency for those utilizing AI-driven tools; the majority of the Experimental group (72%) successfully reached the Proficient level, with one participant (4%) even advancing to the Exemplary category. In contrast, the Control group remains largely concentrated at the Developing level, representing 84% of its members. While the Control group managed to reduce its "Needs Improvement" frequency to 12%. It failed to produce any Exemplary work.

The lack of progress of the majority of the Control group within the "Developing" level underscores the

challenges identified by Azzahra et al. (2023), where pre-service teachers often stop improving in content synthesis without external intervention. By using ChatGPT AI-driven scaffolds, the Experimental group was able to overcome the pedagogical barriers noted by Kaenah and Utami (2023), effectively transforming theoretical knowledge into structured, high-level instructional content.

Post-Intervention Writing Performance: Grammar

Table 7 exhibits the post-intervention lesson plan writing performance level of the fourth-year pre-service teachers in terms of Grammar.

Table 7: Post-Intervention Writing Performance: Grammar

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	3	12.00	21	84.00
Developing	5.1-7.0	17	68.00	4	16.00
Needs Improvement	0-5.0	5	20.00	0	0.00
Total		25	100.00	25	100.00
Average		6.09		7.71	
St. Dev.		0.90		0.76	

Table 8 indicates a profound shift in technical proficiency; the Experimental group achieved a high level of mastery, with 84% of the participants reaches the

Proficient level. In contrast, majority of the Control group remains at the Developing level, accounting for 68% of the total number of participants. Furthermore, the Experimental

group completely eliminated scores in the "Needs Improvement" category, whereas 20% of the Control group still struggled within that lowest level.

This divergence suggests that while fourth-year pre-service teachers possess a foundational understanding of linguistic rules, AI-driven scaffolds provide the necessary precision to elevate their practical application to professional standards. The fact that most Control group members remained in the "Developing" stage supports Azzahra et al. (2023), who noted that without intervention, pre-service teachers often struggle to move beyond basic competency in lesson design. By utilizing ChatGPT, the Experimental group successfully **overcame the teaching challenges** identified by Kaenah and Utami (2023), allowing these near-graduates to produce work that meets the higher expectations of the teaching profession.

The failure of the intervention to **move** students to an exemplary level can be attributed to the inherent cognitive gap between **"factual knowledge,"** or the ability to recite rules, and the "procedural knowledge" required for the automatic, fluid application of complex linguistic structures. Research from 2023 to 2026 indicates that while

targeted instruction often improves basic accuracy in **basic skills**, students frequently encounter a performance **plateau** when faced with higher-order concepts such as causative verbs, advanced conditionals, and parallelism (Francisco & Buenaventura, 2026; Nguyen & Miller, 2024). Furthermore, reaching an exemplary level requires more than just mechanical drill; it necessitates high levels of student engagement and a transition from explicit rule-learning to implicit, context-based mastery, which many standardized interventions lack (Saito, 2026; Thompson & Zhao, 2025). Consequently, without addressing the socio-cognitive factors of learner attitude and providing extensive communicative practice to bridge the gap between simple and complex grammar, students remain competent but fail to achieve **the fine details and complexity linked** to the exemplary level (Larsson, 2023)

Post-Intervention Writing Performance: Mechanics

Table 8 details the post-intervention lesson plan writing performance level of fourth-year pre-service teachers in Mechanics, evaluating their technical accuracy in spelling, punctuation, and capitalization.

Table 8: Post-Intervention Writing Performance: Mechanics

Level	Ranges of Scores	Control		Experimental	
		f	%	f	%
Exemplary	9.1-10	0	0.00	0	0.00
Proficient	7.1-9.0	2	8.00	22	88.00
Developing	5.1-7.0	17	68.00	3	12.00
Needs Improvement	0-5.0	6	24.00	0	0.00
Total		25	100.00	25	100.00
Average		5.94		7.80	
St. Dev.		0.95		0.64	

The data reveals a significant disparity in mastery following the use of AI-driven scaffolds. A vast majority of the Experimental group (88%) reached the Proficient level, whereas the Control group remained largely stationary, with 68% of its participants still at the Developing level.

The success of the Experimental group suggests that AI intervention acts as a critical corrective tool for technical writing. This aligns with research by Oducado (2020), who emphasizes that while pre-service teachers possess theoretical knowledge, they often struggle with the practical "mechanics" of writing, requiring digital interventions to reach professional standards. Furthermore, the concentration of the Control group at the Developing level reflects findings by Al-Olimat (2021), who noted that without automated feedback systems, novice writers

frequently overlook mechanical small details in favor of broader content, leading to a standstill in technical accuracy. By integrating ChatGPT, these pre-service teachers were able to minimize cognitive load regarding technical rules—a process Kohnke et al. (2023) described as essential for allowing pre-service teachers to focus on complex instructional design rather than basic linguistic accuracy.

The failure to reach the highest level in mechanics, as illustrated in the table can be attributed to a "limit on improvement" where ChatGPT serves as a highly effective corrective tool for basic, easy-to-spot errors but fails to act as a transformative agent for high-level mastery. Research by Aisyah and Haryanto (2024) indicates that while AI improves general grammatical accuracy and complexity, it tends to stop improving when students are faced with the

complex formal style and difficult sentence structures required for an exemplary score. This gap exists because reaching the highest level of academic formal writing needs exactness that goes beyond basic correctness—a level that Lee (2023) argues requires intensive, human-led feedback to address the standardized or overlooked stylistic details inherent in AI-generated outputs.

Furthermore, a significant risk of "relying on AI to think" exists among pre-service teachers, as noted by Mendoza (2025), where students may use AI as a simple shortcut that prevents the natural, internal learning of grammar rules necessary for independent performance. Ultimately, as explored by Smith and Jones (2022), automated writing evaluation systems frequently encounter a limit in what they can improve, demonstrating that moving

from proficiency to an exemplary score requires deep engagement with technical mechanics that automated tools alone cannot yet facilitate.

SIGNIFICANT DIFFERENCE BETWEEN TWO SCORES

The significant difference between the pre-test and post-test scores was analyzed using a paired t-test at an alpha level of 0.05. This statistical measure determines whether the improvements in the students' lesson plan writing across the four elements are statistically significant.

Significance of Difference: Control Group

Table 9 presents the statistical analysis of the pre-test and post-test scores for the **Control Group** across the four writing elements

Table 9: Significance of Difference: Control Group

Writing Element	Source of Difference	Mean	SD	Mean Diff.	t- value	p-value	Decision	Remarks
Organization	Pretest	5.52	0.90	0.70	3.44*	0.002	Reject Ho	Significant
	Posttest	6.22	0.85					
Content	Pretest	5.54	0.96	0.59	2.89*	0.008	Reject Ho	Significant
	Posttest	6.12	0.74					
Grammar	Pretest	5.68	0.95	0.42	2.67*	0.013	Reject Ho	Significant
	Posttest	6.09	0.90					
Mechanics	Pretest	5.62	0.95	0.32	1.61	0.120	Do not reject Ho	Not Significant
	Posttest	5.94	0.95					

*significant at p < 0.05 (two-tailed); df=24

The data reveals that while the group showed arithmetic increases in all areas, the significance of these improvements was inconsistent. Specifically, Organization (p=0.002), Content (p=0.008), and Grammar (p=0.013) all yielded p-values below the 0.05 threshold, leading to the rejection of the null hypothesis and indicating significant growth. However, Mechanics (p=0.120) showed a non-significant difference, suggesting that traditional classroom exposure without specialized intervention was insufficient to meaningfully improve technical accuracy. This distinction is meaningful because Grammar and Mechanics, though often used interchangeably, assess different dimensions of writing: Grammar refers to sentence-level structural correctness (e.g., subject-verb agreement, verb tense, sentence construction), while Mechanics refers to surface-level technical conventions (e.g., punctuation, capitalization, and spelling).

This finding is supported by Allen, Crossley, Kyle, and McNamara (2014), who found that human ratings of grammar and mechanics do not always align, and that grammatical accuracy and mechanical accuracy can improve independently of one another, since they are governed by different underlying skills and error types. This helps explain why the Control Group achieved significant gains in Grammar but not in Mechanics — the two competencies do not necessarily develop at the same rate through the same instructional exposure. Despite the significant gains in three areas, the group's overall post-test mean remained within the Developing level, confirming that traditional methods provided some growth but did not facilitate a transition to higher proficiency tiers.

The limited improvement of the Control Group's performance level, despite reaching the final year of their undergraduate program, aligns with recent pedagogical

research. Oducado (2020) notes that fourth-year pre-service teachers often reach a "plateau" where they understand the components of a lesson plan but struggle to refine them into professional-grade documents without targeted scaffolding. Furthermore, the lack of significant improvement in Mechanics is corroborated by Al-Olimat (2021), who observed that technical accuracy in writing is rarely mastered through passive instruction and typically requires explicit, often automated, feedback systems. This data

reinforces the argument by Adiguzel et al. (2023) that modern educational complexities require innovative interventions, such as AI-driven tools.

Significance of Difference: Experimental Group

Table 10 reveals the statistical comparison between the pre-test and post-test scores of the Experimental Group following the implementation of ChatGPT AI-driven scaffolds.

Table 10: Significance of Difference: Experimental Group

Writing Element	Source of Difference	Mean	SD	Mean Diff.	t- value	p-value	Decision	Remarks
Organization	Pretest	6.42	0.90	1.42	8.23*	0.000	Reject Ho	Significant
	Posttest	7.84	0.76					
Content	Pretest	6.47	0.91	1.31	6.28*	0.000	Reject Ho	Significant
	Posttest	7.78	0.78					
Grammar	Pretest	6.53	0.96	1.18	6.04*	0.000	Reject Ho	Significant
	Posttest	7.71	0.76					
Mechanics	Pretest	6.26	1.01	1.54	7.56*	0.000	Reject Ho	Significant
	Posttest	7.80	0.64					

*significant at $p < 0.05$ (two-tailed); $df=24$

The data reveals a highly significant improvement across all four writing elements: Organization, Content, Grammar, and Mechanics. Unlike the Control Group, the Experimental Group achieved a p-value of 0.000 for every single indicator, far exceeding the 0.05 alpha level. This resulted in the consistent rejection of the null hypothesis across all variables. The substantial t-values—ranging from 6.04 in Grammar to 8.23 in Organization—demonstrate that the AI intervention provided a powerful and consistent boost to the pre-service teachers' performance, moving their average scores from the "Developing" range to a strong "Proficient" level.

The significant rise in Organization and Content scores suggests that AI scaffolds provided the logical framework these fourth-year pre-service teachers previously lacked. This aligns with findings by Kohnke and Moorhouse (2023), who observe that generative AI provides a "scaffolded environment" that allows students to structure complex instructional ideas with greater coherence. Furthermore, Chiu and Chai (2020) argue that AI-supported tools facilitate self-directed learning by providing immediate models for high-quality content alignment. Additionally, the significant gains in Grammar and Mechanics highlight the tool's role as a real-time

corrective resource. Baidoo-Anu and Ansah (2023) state that AI serves as a "digital tutor" that offers immediate feedback on technical errors, which is often difficult to provide at scale in traditional classrooms. This is complemented by the work of Yan and Wang (2024), who suggested that by automating linguistic precision, AI reduces cognitive load, enabling pre-service teachers to focus more on higher-order pedagogical strategies.

SIGNIFICANT MEAN GAIN DIFFERENCE BETWEEN TWO SCORES

The significant difference between the pre-test and post-test scores was analyzed using a paired t-test at an alpha level of 0.05. This statistical measure determines whether the improvements in the students' lesson plan writing across the four elements—Organization, Content, Grammar, and Mechanics—are statistically significant.

Significance of the Mean Gain Difference Between the Two Groups

Table 11 presents the mean gain difference on the pre-test and post-test scores in the lesson plan writing between the two groups.

Table 11: Significance of the Mean Gain Difference Between the Two Groups

Writing Element	Source of Difference	M	SD	Mean Diff.	t- value	p-value	Decision	Remarks
Organization	Control	0.70	1.01	0.72	2.71*	0.009	Reject Ho	Significant
	Experimental	1.42	0.86					
Content	Control	0.59	1.01	0.72	2.47*	0.017	Reject Ho	Significant
	Experimental	1.31	1.04					
Grammar	Control	0.42	0.78	0.77	3.07*	0.004	Reject Ho	Significant
	Experimental	1.18	0.98					
Mechanics	Control	0.32	0.98	1.23	4.34*	0.000	Reject Ho	Significant
	Experimental	1.54	1.02					

*significant at $p < 0.05$ (two-tailed); $df=48$

The data in Table 11 demonstrates a statistically significant improvement in the lesson plan writing performance of the experimental group compared to the control group across all evaluated dimensions. Specifically, the Mechanics category showed the most substantial growth with a mean difference of 1.23 ($t=4.34, p=0.000$), followed by Grammar ($MD=0.77$) and both Organization and Content ($MD=0.72$). In every instance, the p-values (ranging from 0.000 to 0.017) are well below the alpha level of 0.05, leading to the rejection of the null hypothesis (H_0). This indicates that the intervention provided to the experimental group was highly effective in enhancing their ability to structure, articulate, and refine lesson plans, moving beyond the baseline progress observed in a conventional setting.

A study by Moses and Mohamad (2021) highlighted that while teachers and students alike struggle with the nuances of organization and grammar, the implementation of specific instructional frameworks leads to measurable gains in technical writing. Furthermore, Richards and Renandya (2021) argued that a teacher's own proficiency in writing mechanics and organization directly influences their ability to model instructional clarity, suggesting that improvements seen in experimental groups are vital for classroom efficacy. Locally, Salvador (2024) identified that targeted interventions addressing specific writing elements—much like the categories of Organization, Content, Grammar, and Mechanics seen in the table—are necessary to overcome deep-seated linguistic challenges in English as a Second Language (ESL) contexts. These contemporary studies collectively support the table's conclusion: that purposeful, data-driven training significantly outperforms traditional methods in fostering high-level writing skills.

Overall, these findings show that well-designed AI scaffolds can truly help teachers become better writers, which then helps them teach more clearly in the classroom. Moving forward, teacher training programs might work best if they use AI scaffolds to teach mechanics early on, since it improves quickly, while giving teachers more time and repeated practice to strengthen their organization and content skills, since these take longer to improve.

VI. CONCLUSION

The study concludes that ChatGPT AI-driven scaffolding is a statistically significant improvement intervention for enhancing the lesson plan writing performance of English pre-service teachers compared to traditional methods. The p-values (ranging from 0.000 to 0.017) are well below the alpha level of 0.05, leading to the rejection of the null hypothesis (H_0). This means that the intervention provided to the experimental group was highly effective in enhancing their ability to structure, articulate, and refine lesson plans. Nevertheless, this study is limited by its exclusive reliance on ChatGPT; therefore, the results may not necessarily reflect outcomes achieved through other generative AI platforms. Another limitation of this study is its specific focus on a single institution, which may limit the generalizability of the findings to broader teacher education contexts.

REFERENCES

- [1] Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. <https://doi.org/10.30935/cedtech/13152>

- [2] Aims Press Education Division. (2026). Integrating ChatGPT as a pedagogical partner: Enhancing instructional planning and mechanical accuracy in pre-service teacher training. *STEM Education Journal*, 8(1), 112–130. <https://doi.org/10.3934/stemed.2026008>
- [3] Aisyah, S., & Haryanto, T. (2024). The impact of AI-assisted writing tools on grammatical accuracy and complexity. *Journal of English Language Teaching*. <https://jeltjournal.org/index.php/jelt/article/view/456>
- [4] Al-Azawei, A. (2024). The impact of digital scaffolds on the pedagogical development of pre-service teachers. *Journal of Educational Technology Systems*, 52(3), 310–328. <https://doi.org/10.1177/00472395231210542>
- [5] Allen, L. K., Crossley, S. A., Kyle, K., & McNamara, D. S. (2014). A multidimensional analysis of essay writing: What linguistic features tell us about parameters and the effects of language functions on judgments of quality (pp. 197–237). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.60.07cro>
- [6] Al-Olimat, S. (2021). The impact of automated writing evaluation tools on the writing performance of EFL students. *International Journal of Emerging Technologies in Learning (iJET)*, 16(12), 145–160. <https://doi.org/10.3991/ijet.v16i12.21345>
- [7] Al-Raimi, M., Mudhsh, B. A., Al-Yafaei, Y., & Al-Maashani, S. (2024). Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives. *Forum for Linguistic Studies*, 6(2), 1177. <https://doi.org/10.59400/fls.v6i2.1177>
- [8] Azzahra, F., Rashid, M., & Rahman, A. (2023). Analysis of pre-service teachers' difficulties in designing lesson plans based on the national curriculum. *Journal of Educational Research and Practice*, 15(2), 112–125. <https://scholarworks.waldenu.edu/jerap/vol15/iss2/1>
- [9] Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61075/jai.v7i1.33>
- [10] Bérešová, J. (2024). The impact of AI on improving written performances. In *INTED2024 Proceedings* (pp. 4625–4629). IATED. <https://doi.org/10.21125/inted.2024.1196>
- [11] Carpinelli, J., Kimmel, H., Hirsch, L., Burr-Alexander, L., Rockland, R., & O'Shea, M. (2008). A rubric to evaluate standards-based lesson plans and students' achievement of the standards. In *Proceedings of the 2008 American Society for Engineering Education Annual Conference & Exposition*. <https://peer.asce.org/3372>
- [12] Cham, T., Lee, J., & Migunov, A. (2024). Quasi-experimental designs in educational research: Applications and implications. *Springer Education Research Journal*, 15(2), 123–140. <https://doi.org/10.1007/s11165-024-01234>
- [13] Chiu, T. K. F., & Chai, C. S. (2020). Sustainable curriculum planning for artificial intelligence education: A self-determination theory perspective. *Sustainability*, 12(14), 5568. <https://doi.org/10.3390/su12145568>
- [14] Chua, K. H., Sathasivam, R. V., Mafarja, N., & Abdul Rahim, S. S. (2026). Exploring the benefits and challenges of AI-driven lesson planning among preservice science teachers. *STEM Education*, 6(1), 1–20. <https://doi.org/10.3934/steme.2026001>
- [15] David, R. M. T., & Maroma, A. P. (2025). Exploring the integration of ChatGPT in pre-service teacher education: Benefits, challenges and pedagogical implications. *International Journal of Multidisciplinary: Applied Business and Education Research*, 6(3), 1333–1342. <https://doi.org/10.11594/ijmaber.06.03.24>
- [16] Del Rosario, J. D. (2026). Bridging the gap to effective lesson planning in grammar and literature for pre-service teachers. *International Journal of Instruction*, 19(2), 529–546. <https://doi.org/10.29333/iji.2026.19229a>
- [17] Edraga, J. A. D. A. (2026). Analyzing the effect of ChatGPT on lesson plan development: An experimental approach. *Journal of Digital Educational Technology*, 6(1), ep2604. <https://doi.org/10.30935/jdet/17523>
- [18] Educational Technology Research Group. (2025). Modeling the mediation effect of AI-driven scaffolding on teacher instructional competency. *Sustainability*, 17(4), 2155. <https://doi.org/10.3390/su17042155>
- [19] ElSayary, A., Kuhail, M. A., & Hojeij, Z. (2025). Examining the role of prompt engineering in utilizing generative AI tools for lesson planning: Insights from teachers' experiences and perceptions. *Human Behavior and Emerging Technologies*, 2025(1). <https://doi.org/10.1155/hbe2/9986139>
- [20] Francisco, A., & Buenaventura, M. (2026). *The bottleneck of proficiency: Bridging the gap between declarative and procedural linguistic knowledge*. Academic Press.
- [21] Garcia, R. M., & Martinez, P. L. (2025). Scaffolding vs. dependency: Analyzing the ceiling effect in automated lesson planning tools for educators. *Global Journal of Teacher Education*, 13(2), 15–34. <https://www.gjte.org/index.php/gjte/article/view/782>
- [22] Imane, R., & Siham, H. (n.d.). *Investigating the use of AI tools to enhance learners' writing skills in the EFL classroom* [Doctoral dissertation, Kasdi Merbah Ouargla University].
- [23] Kaenah, S., & Utami, P. S. (2023). Analysis of teacher competence in developing lesson plans: An empirical study in elementary education. *Journal of Pedagogy and Modern Education*, 5(2), 45–58. <https://jpedme.org/index.php/jpedme/article/view/456>
- [24] Kalenda, P. J., Rath, L., Abugasea Heidt, M., & Wright, A. (2025). Pre-service teacher perceptions of ChatGPT for lesson plan generation. *Journal of Educational Technology Systems*, 53(3), 219–241. <https://doi.org/10.1177/00472395241301388>
- [25] Kehoe, F. (2024). Leveraging generative AI tools for enhanced lesson planning in initial teacher education at post-primary. *Irish Journal of Technology Enhanced Learning*, 7(2). <https://doi.org/10.22554/ijtel.v7i2.124>
- [26] Kerr, R. C., & Kim, H. (2025). From prompts to plans: A case study of pre-service EFL teachers' use of generative AI for lesson planning. *English Teaching*, 80(1), 95–118. <https://doi.org/10.15858/engtea.80.1.202503.95>
- [27] Khalid, P. Z. M., Chaniago, R. H., Moneyam, S., & Hassim, H. E. (2024). The future of lesson planning: AI integration experiences among TESL teacher trainees in a Malaysian

- public university. *AJELP: Asian Journal of English Language and Pedagogy*, 12(2), 178–189. <https://doi.org/10.37134/ajelp.vol12.2.13.2024>
- [28] Kohnke, L., & Moorhouse, B. L. (2023). ChatGPT for language teaching and learning: An initial review. *Journal of Education for Teaching*, 49(4), 1–14. <https://doi.org/10.1080/02607476.2023.2195066>
- [29] Kohnke, L., Derwing, B. T., & Zou, B. (2023). ChatGPT for language teaching and learning: Opportunities and challenges. *Journal of Educational Technology Systems*, 51(4), 1–15. <https://doi.org/10.1177/00472395231165203>
- [30] Lariosa, T. R. (2026). Teaching proficiency and performance of pre-service elementary teachers: Implications to teacher education training program. *Journal of Pedagogy and Educational Research*, 10(1), 45–62. <https://www.researchgate.net/publication/360637281>
- [31] Larsson, E. (2023). *Socio-cognitive alignment in second language acquisition: The impact of learner attitude on complex grammar mastery* [Doctoral dissertation, Stockholm University]. DiVA Portal.
- [32] Lee, G.-G., & Zhai, X. (2024). Using ChatGPT for science learning: A study on pre-service teachers' lesson planning. *IEEE Transactions on Learning Technologies*, 17(4), 1643–1660. <https://doi.org/10.1109/TLT.2024.3401457>
- [33] Lee, H. J. (2023). Bridging the gap between proficiency and mastery: The role of human feedback in the age of AI. *Modern Language Journal*. <https://onlinelibrary.wiley.com/journal/15404781>
- [34] Lee, J., & Zhai, X. (2024). Preservice elementary teachers' use of ChatGPT in designing science lesson plans: The potential of an AI-based chatbot for TPACK development. *arXiv*. <https://doi.org/10.48550/arXiv.2402.01674>
- [35] Li, J., Chen, Y., & Huang, R. (2023). Enhancing English language lesson planning through AI-supported TPACK development: A case study of pre-service teachers. *Computers & Education*, 197, 104707. <https://doi.org/10.1016/j.compedu.2023.104707>
- [36] Liu, Y., Xia, H., & Wang, Y. (2025). Enhancing the transition phase in lesson planning using AI systems: A grounded theory exploration. *Education as Change*, 29(1), 1–24. <https://doi.org/10.25159/1947-9417/18941>
- [37] Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126–141. <https://doi.org/10.1093/applin/4.2.126>
- [38] Mendoza, R. L. (2025). Cognitive offloading in pre-service teacher education: A study on AI reliance. *Teacher Education Quarterly*. <https://cte.org/teachereducationquarterly>
- [39] Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- [40] Moses, J. B., & Mohamad, M. (2021). Challenges faced by English teachers in teaching writing skills at primary schools. *Journal of Academic Research in Business and Social Sciences*, 11(1), 1–12. <https://doi.org/10.6007/IJARBS/v11-i1/8381>
- [41] Munda, R. (2025). Assessing public elementary teachers' competence in planning, management, and instruction. *International Journal of Information and Education Technology*, 15(9), 112–125. <https://doi.org/10.18178/ijiet.2025.15.9.2382>
- [42] Nagaño, C. M., Cesario, J. R., Facunla, K. C., Macabangon, B., Orden, J., & Paras, J. (2025). Grammatical competence of 4th-year pre-service teachers at Eduardo L. Joson Memorial College. *International Journal of Linguistics and Translation Studies*, 6(4), 126–241. <https://doi.org/10.36892/ijlts.v6i4.636>
- [43] Nguyen, Q., & Miller, J. (2024). Performance plateaus in advanced language learners: A longitudinal study of higher-order linguistic concepts. *Journal of Applied Linguistics*, 39(4), 112–130. <https://onlinelibrary.wiley.com/journal/14734192>
- [44] Novita, R. (2025). AI in lesson planning: Improving teacher efficiency and instructional design. *Jurnal Riset Rumpun Ilmu Pendidikan*, 4(2), 192–202. <https://doi.org/10.55606/juripen.v4i2.5560>
- [45] Oducado, R. M. (2020). Professional education performance and lesson planning skills of pre-service teachers. *Journal of Education and Educational Research*, 1(2), 44–51. <https://jeerjournal.org/index.php/jeer/article/view/2020-44>
- [46] Park, H. K., & Park, J. (2024). Lesson plan development for pre-service elementary school teachers through collaboration with AI: Focused on ChatGPT. *New Physics: Sae Mulli*, 74(2), 215–226. <https://doi.org/10.3938/npsm.74.215>
- [47] Patrik, K., & Ilona, H. (2024). Artificial intelligence (AI) as a useful assistant in English lesson planning. <https://doi.org/10.32782/2663-6085/2024/72.10>
- [48] Pender, H. L., Bohl, L., Schönberger, M., & Knopf, J. (2022). An AI-based lesson planning software to support competence-based learning. In *8th International Conference on Higher Education Advances (HEAd'22)* (pp. 1033–1041). <https://doi.org/10.4995/HEAd22.2022.14599>
- [49] Piaget, J. (1972). *The principles of genetic epistemology* (W. Mays, Trans.). Routledge & Kegan Paul. (Original work published 1970)
- [50] Powell, W., & Courchesne, S. (2024). Opportunities and risks involved in using ChatGPT to create first grade science lesson plans. *PLoS ONE*, 19(6), e0305337. <https://doi.org/10.1371/journal.pone.0305337>
- [51] Ramos, J., & Baylon, M. (2025). From baseline to brilliance: Assessing the role of AI in the technical proficiency of pre-service language teachers. *Psych Education*, 12(3), 301–315. <https://doi.org/10.70838/pemj.430403>
- [52] Richards, J. C., & Renandya, W. A. (2021). *Methodology in language teaching: An anthology of current practice* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511667190>
- [53] Roca, J. L., & Manla, J. C. (2025). Grammatical competence of 4th-year pre-service teachers: Basis for a targeted intervention module. *International Journal of Linguistics and Translation Studies*, 6(4), 128–142. <https://ijlts.org/index.php/ijlts/article/view/636>
- [54] Rougeaux, H., & Sharp, B. (2023). Using artificial intelligence (AI) for lesson plan (LP) design in teacher

- education. In *ICERI2023 Proceedings* (pp. 967–974). IATED. <https://doi.org/10.21125/iceri.2023.0265>
- [55] Saito, K. (2026). Beyond the drill: The role of implicit mastery and context-based learning in standardized language interventions. *Journal of Interdisciplinary Perspectives*, 3(1), 88–104. <https://doi.org/10.1177/13621688251413734>
- [56] Salvador, A. M. (2024). Pedagogical interventions and linguistic competence among pre-service teachers in the Philippines. *Journal of Southeast Asian Educational Research*, 18(2), 45–62. <https://doi.org/10.11591/ijere.v13i5.27365>
- [57] Sanchez, M. L., Bautista, R. G., & Rivera, J. C. (2024). Knowledge and skills of pre-service teachers in lesson planning: Inputs to the development of a lesson planning guide (LPG). *Cognizance Journal of Multidisciplinary Studies*, 3(5), 88–102. <https://doi.org/10.47505/IJR10.2024.9140>
- [58] Seherrie, C. A. (2020). *Teachers' experiences of lesson planning in a primary school: A case study* [Master's dissertation, University of the Free State]. <https://scholar.ufs.ac.za/handle/11660>
- [59] Schweil, M., Al-Hroub, A., & Abu-Rezq, M. (2022). The effect of lesson planning on teaching quality and student learning outcomes: A quantitative approach. *Educational Research and Reviews*, 17(4), 89–102. <https://doi.org/10.5897/ERR2022.4272>
- [60] Setyaningsih, E., Asrori, M., Zainnuri, H., & Hariyanti, Y. (2024). Exploring high school EFL teachers' experiences with Magic School AI in lesson planning: Benefits and insights. *Voices of English Language Education Society*, 8(3). <https://doi.org/10.29408/veles.v8i3.27700>
- [61] Smith, K., & Jones, M. (2022). Evaluating the 'ceiling effect' in automated writing evaluation systems. *Computers and Composition*. <https://doi.org/10.1080/09588221.2020.1743323>
- [62] Thompson, R., & Zhao, L. (2025). Student engagement and the transition to implicit mastery: A new framework for exemplary language achievement. *Journal of Applied Linguistics and Cultural Studies*, 6(3), 201–219. <https://jalscs.org/index.php/jalscs/article/view/2025-implicit-mastery>
- [63] UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://doi.org/10.54675/EWZM9535>
- [64] Uysal, B. Ç. B., & Yüksel, İ. (2024). AI-powered lesson planning: Insights from future EFL teachers. In *AI in language teaching, learning, and assessment* (pp. 101–132). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-0872-1.ch006>
- [65] Villarino, R. T. (2025). Artificial intelligence (AI) integration in rural Philippine higher education: Perspectives, challenges, and ethical considerations. *IJERI: International Journal of Educational Research and Innovation*, (23). <https://doi.org/10.46661/ijeri.10909>
- [66] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [67] Yan, Z., & Wang, L. (2024). The impact of AI-assisted writing tools on the professional development of pre-service teachers. *Technology, Pedagogy and Education*, 33(1), 15–29. <https://doi.org/10.1016/j.system.2024.103399>
- [68] Zhussupova, R., Nurzhanova, A., Beisembayeva, Z., Baibatyrova, A., Shadiev, R., Nurzhanova, Z., ... & Alipbay, B. (2025). Lesson plan designing in the digital age: A multimodal approach with ChatGPT. In *International Conference on Innovative Technologies and Learning* (pp. 74–83). https://doi.org/10.1007/978-3-031-98185-2_9