Pedagogical Research
2025, 10(4), em0251
e-ISSN: 2468-4929
https://www.pedagogicalresearch.com

MODESTUM

Research Article **OPEN ACCESS**

A comparative study of Google Gemini and ChatGPT in enhancing english language learning for EFL learners: A case study of the english research writing course

Pongpatchara Kawinkoonlasate [1*]

[1] Faculty of Liberal Arts, Huachiew Chalermprakiet University, THAILAND
***Corresponding Author:** fai_pimol@hotmail.com

Citation: Kawinkoonlasate, P. (2025). A comparative study of Google Gemini and ChatGPT in enhancing english language learning for EFL learners: A case study of the english research writing course. *Pedagogical Research, 10*(4), em0251. https://doi.org/10.29333/pr/17670

ARTICLE INFO

Received: 14 Oct 2025

Accepted: 07 Dec 2025

ABSTRACT

This study compared Google Gemini and ChatGPT's effectiveness for enhancing the academic writing skills of English as a Foreign Language (EFL) learners in eight-week research writing course in two Thai universities. Using a quasi-experimental approach, 80 third-year learners of a private university (HCU) and a public university (NRRU) were assigned to use either ChatGPT or Google Gemini for writing tasks. Pre- and post-intervention writing assessments, Technology Acceptance Model-inspired Likert-scale questionnaires, and semi-structured interviews were employed to measure writing ability gains, learner attitudes, and learning engagement. The results indicate that both tools effectively improved linguistic accuracy and essay structure, and Gemini was found to work better than ChatGPT in multimodal feedback and source integration, particularly among rural learners ($p < .05$, Cohen's $d = 0.62$). However, overdependence and plagiarism risk were cited as issues. Qualitative findings suggest increased learner confidence and motivation, but also require adjustment of implementation in rural areas because of cultural and technological constraints. This underscores AI's promise in EFL writing instruction, emphasizing instructor guidance and AI literacy in Thai contexts. Pedagogical implications include hybrid AI models for low-resource settings, with calls for further longitudinal research.

Keywords: Google Gemini, ChatGPT, EFL learners, academic writing, research writing course

BACKGROUND OF THE STUDY

The rapid advancement of Artificial Intelligence (AI) has revolutionized educational practices, particularly in English as a Foreign Language (EFL) instruction, by introducing innovative approaches to language learning. AI-based applications have emerged as powerful tools for fostering language skills, promoting learner autonomy, and enhancing engagement in diverse educational contexts (Baskara, 2025). Among these tools, generative AI platforms such as ChatGPT, developed by OpenAI, and Google Gemini, created by Google DeepMind, stand out for their ability to provide real-time support, advanced language modeling, and interactive dialogue, which are critical for supporting EFL learners in developing academic writing proficiency (Dwivedi et al., 2023; Zawacki-Richter & Jung, 2023). This study examines the comparative effectiveness of ChatGPT and Google Gemini in enhancing the academic writing skills of EFL learners within a research writing course, with a particular focus on third-year learners in Thai university settings. The analysis centers on four key dimensions:

1) Providing personalized feedback,
2) Improving linguistic accuracy,
3) Fostering motivation, and
4) Addressing challenges such as plagiarism, authenticity, and technological barriers.

Integrating AI into EFL education has transformed traditional pedagogical approaches by offering scalable, personalized, and interactive learning experiences. AI tools utilize natural language processing (NLP) and machine learning to provide instant feedback, simulate real-world communication, and support self-directed learning, which are particularly beneficial for English as a Foreign Language (EFL) learners transitioning from basic to advanced language skills (Baskara, 2025). These tools address common challenges in EFL instruction, such as limited teacher availability for individualized feedback, varying proficiency levels among learners, and the need for engaging, contextually relevant materials. By offering tailored support, AI platforms empower learners to improve their writing, speaking, and critical thinking skills, thereby fostering greater autonomy and confidence (Zawacki-Richter & Jung, 2023).

In Thai EFL contexts, where English proficiency is often a prerequisite for academic and professional success, the demand for innovative tools to support language development is particularly pronounced. Third-year university learners, who are navigating the transition from foundational English to academic writing mastery, face unique challenges, including producing research papers that require advanced linguistic accuracy, structural coherence, and adherence to academic conventions (Phakiti & Paltridge, 2015). AI tools like ChatGPT and Google Gemini offer promising solutions by providing real-time assistance and fostering engagement, but their comparative effectiveness in this specific context remains underexplored. ChatGPT, built on OpenAI's GPT architecture, has gained widespread recognition for its conversational fluency and ability to deliver context-specific feedback, making it a valuable tool for EFL writing instruction. In research writing courses, ChatGPT supports learners by identifying grammatical errors, suggesting vocabulary enhancements, and providing structural guidance (e.g., improving thesis statements or paragraph coherence). Its interactive dialogue feature allows learners to ask follow-up questions, enabling iterative refinement of their drafts (Bok & Cho, 2024). For instance, a learner might submit a draft abstract to ChatGPT and receive feedback on clarity, academic tone, and source integration, followed by tailored suggestions for improvement. Studies have shown that ChatGPT enhances EFL learners' writing performance by reducing cognitive load and providing immediate, non-judgmental feedback, which is particularly valuable in large classes where individualized teacher attention is limited (Yan, 2023).

Beyond feedback, ChatGPT supports writing fluency by generating sample texts, such as essay outlines or model paragraphs, which serve as scaffolds for EFL learners. Song and Song (2023) found that EFL learners using ChatGPT in academic writing tasks demonstrated significant improvements in grammar, vocabulary diversity, and coherence, particularly in research-oriented writing. However, challenges include the risk of overreliance, which may lead to homogenized writing styles or reduced critical thinking, as well as the potential for plagiarism if learners submit AI-generated content as their own (Fitria, 2023). Whereas Google Gemini, a more recent AI platform developed by Google DeepMind, integrates multimodal capabilities and leverages Google's expansive ecosystem to provide context-rich, visually enhanced learning experiences. Unlike ChatGPT, which relies primarily on text-based interactions, Gemini's ability to process and generate multimedia content (e.g., images, videos, and web-based data) makes it uniquely suited for supporting vocabulary acquisition, contextual understanding, and content relevance in academic writing (Park & Gupta, 2024). In research writing course, Gemini can assist learners by suggesting up-to-date sources, verifying factual claims, and recommending field-specific terminology, which is critical for producing credible research papers. For example, Pratiwi et al. (2024) found that Gemini improved the collocation proficiency of EFL learners in a Paragraph Writing course, suggesting its potential for enhancing linguistic precision in academic contexts. Moreover, Gemini's real-time web access allows it to provide feedback grounded in current academic resources, which is particularly valuable for research writing tasks requiring recent data or citations.

However, its feedback may be less conversational than ChatGPT's, potentially limiting iterative refinement, and its performance in non-English languages or culturally specific contexts may be less robust (Sallam et al., 2024). Additionally, Gemini's reliance on web-based data introduces variability in feedback quality, depending on the sources accessed, which may require learners to critically evaluate outputs. In Thai EFL settings, particularly at the university level, the integration of AI tools aligns with the national push to enhance English proficiency as part of the Thailand 4.0 initiative, which emphasizes digital innovation in education (Chutiphongdech & Vongvilai, 2021). Third-year university learners, often transitioning from general English courses to discipline-specific academic writing, require support in mastering complex linguistic structures, academic conventions, and critical thinking skills. ChatGPT's detailed grammar and vocabulary feedback addresses these needs by helping learners refine their research papers, while Gemini's visual aids and web-integrated content enhance vocabulary acquisition and contextual understanding. For example, a Thai EFL learner writing a literature review might use ChatGPT to improve sentence-level accuracy and Gemini to identify recent studies, ensuring both linguistic and content quality. In speaking tasks, which are often integrated into research writing courses through presentations or discussions, ChatGPT's conversational simulation supports fluency and confidence, while Gemini's multimedia capabilities could enhance pronunciation practice through audio-visual feedback (Park & Gupta, 2024). Both tools foster engagement by making learning interactive and accessible, but their effectiveness in Thai classrooms depends on addressing challenges such as technological access, AI accuracy, and cultural relevance.

The reason for choosing ChatGPT and Google Gemini over other platforms like Grammarly, Claude, or Llama is due to their higher versatility and presence in the Thai context. Grammarly and such platforms are more focused on grammar and style checks without having the conversation and research aspects of ChatGPT and Gemini (Aydin & Altınay, 2025). Claude and Llama are bound in Thailand because they have limited access and lack real-time data and multimodal functionalities; therefore, they are less suitable for this research study's purposes (Perera & Lankathilaka, 2023). The use of ChatGPT and Gemini is also supported by the fact that they address the specific needs of Thai EFL learners, who require help adapting to higher-level academic writing, and can close research gaps since limited research has compared their effectiveness in Thai EFL settings (Kotmungkun et al., 2024; Meniado, 2023).

Despite their potential, both ChatGPT and Google Gemini face challenges in EFL instruction. Plagiarism remains a significant concern, as learners may misuse AI-generated content, undermining academic integrity (Fitria, 2023). Authenticity issues arise when learners over-rely on AI tools, potentially stifling their unique voice or critical thinking skills (Barrot, 2023). Technological barriers, such as limited internet access or digital literacy among Thai learners, particularly in rural areas, may also hinder effective AI integration (Chutiphongdech & Vongvilai, 2021). Furthermore, while ChatGPT has been extensively studied in EFL contexts, empirical evidence on Gemini's application in academic writing, especially in Thai settings, is limited, necessitating further research to validate its efficacy (Pratiwi et al., 2024). The lack of comparative studies analyzing ChatGPT and Gemini in real classroom settings, particularly for research writing courses, underscores the need for this study. Understanding how these tools influence learner performance, participation, and motivation is critical for educators seeking to leverage AI effectively. This study

aims to fill this gap by examining the comparative impact of ChatGPT and Gemini on Thai EFL learners' academic writing skills, with a focus on feedback quality, linguistic accuracy, motivation, and ethical considerations.

Few studies compare their effectiveness in fostering academic writing skills among third-year Thai EFL learners, who require targeted support to master research writing. This gap risks suboptimal tool use, potentially hindering progress in writing, vocabulary, and grammar (Kotmungkun et al., 2024). Challenges such as limited teacher AI literacy, accuracy issues, technological barriers, and plagiarism concerns further complicate integration (Baskara, 2025). Therefore, this comparative study is significant for several reasons. First, it addresses the growing interest in AI-driven EFL instruction in Thai universities, where academic writing proficiency is a key determinant of learner success. Second, it provides insights into the strengths and limitations of ChatGPT and Gemini, informing pedagogical strategies for their integration. Third, it tackles critical challenges, such as plagiarism and technological barriers, offering practical recommendations for ethical and effective AI use. By focusing on third-year Thai EFL learners in research writing course, the study contributes to the global discourse on AI in education while addressing context-specific needs in Thailand.

Research Objectives

The study aims to:

1. Evaluate the effectiveness of ChatGPT and Google Gemini in providing personalized feedback to improve the academic writing skills of third-year Thai EFL learners in research writing course across two different institutions.
2. Examine the role of ChatGPT and Google Gemini in fostering motivation and engagement among third-year Thai EFL learners during research writing tasks across two different institutions.
3. Identify and analyze the challenges (e.g., plagiarism, authenticity, and technological barriers) associated with using ChatGPT and Google Gemini in Thai EFL research writing instruction across two different institutions.

Research Questions

To achieve the above objectives, the study addresses the following research questions:

RQ1 How effective are ChatGPT and Google Gemini in providing personalized feedback to enhance the academic writing skills of third-year Thai EFL learners in research writing course across two different institutions?

RQ2 How do ChatGPT and Google Gemini influence the motivation and engagement levels of third-year Thai EFL learners in a research writing course across two different institutions?

RQ3 What challenges (e.g., plagiarism, authenticity, and technological barriers) arise when using ChatGPT and Google Gemini in Thai EFL research writing instruction between two different institutions?

Conceptual Framework

The framework posits that ChatGPT and Google Gemini, as independent variables, influence academic writing proficiency through their distinct capabilities. ChatGPT's conversational feedback fosters iterative revisions and linguistic accuracy, while Gemini's multimodal and web-integrated features enhance content relevance and vocabulary acquisition. These processes are mediated by the quality of feedback, linguistic improvements, and motivational impact, but are constrained by challenges like plagiarism, authenticity, and technological barriers. The Thai EFL context, particularly for third-year learners, moderates these relationships due to cultural, linguistic, and infrastructural factors. The framework guides the study by linking AI tools to measurable outcomes in writing proficiency, while addressing ethical and practical challenges (See **Figure 1**).

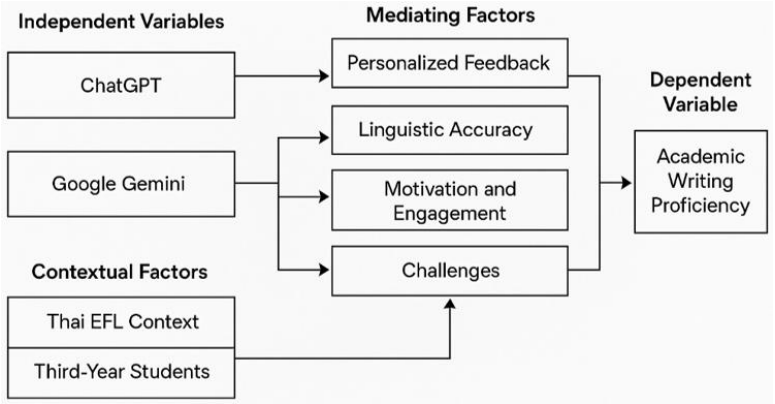

Figure 1. Conceptual framework illustrating the conditional influence of ChatGPT and Google Gemini on EFL academic writing proficiency, mediated by feedback and motivation, moderated by institutional and learner factors, and constrained by ethical and infrastructural barriers (Adapted from Barrot, 2024; Davis, 1989; Phakiti & Paltridge, 2015)

LITERATURE REVIEW

Artificial intelligence, particularly in the form of natural language processing, has brought about a considerable change to language education for English as a Foreign Language (EFL) learners, providing a new arsenal of tools for teaching and learning.

This paper examines the development and adaptation of AI in language education, highlighting the impact of generative AI models, such as large language models (LLMs), and the transition from traditional language tools, like Grammarly, to cutting-edge conversational AIs, including ChatGPT and Google Gemini. Combining these analyses, this review reveals AI's ability to customize and enhance the language learning experiences of EFL learners, with a focus on academic writing. Platforms such as ChatGPT (created by OpenAI) and Google Gemini have broadened the horizons of AI in EFL teaching and learning by providing interactive, conversational, and multimodal learning opportunities. These developments fit well with the increasing calls for more personal and learner-centred methods to language teaching in EFL contexts, not the least of which in programs focusing on academic writing, where learners need scaffolded support in structure, argumentation, and critical thinking (Barrot, 2023; Baskara, 2025). Addintionally, recent research underlines the role of AI in EFL, such as the meta-analysis of Jantakoon et al. (2025), who assessed 19 experimental studies between 2017 and 2024 and found overall effects for AI-enhanced instruction on speaking proficiency of SMD = 1.033, and positive directions in listening, though suggesting a need for tailored AI applications when learners' contexts are heterogeneous. This complements findings on writing, suggesting AI's multimodal potential extends beyond text to oral skills, informing hybrid models in our study.

Overview of Google Gemini and ChatGPT in EFL Writing

Taken together, the development of AI writing tools captures a paradigm shift from surface editing to deep and cognitive contextual support through LLMs. ChatGPT, through OpenAI's GPT architecture, is good at conversational fluency and iterative revision. Thus, it allows error correction, enhancing vocabulary and structural feedback in real-time (Bok & Cho, 2024; Song & Song, 2023). Experimental studies report significant pre-post gain in grammar, coherence, and learner autonomy. In this regard, significant pre-post gains were reported in grammar ($d = 0.68$), coherence ($d = 0.55$), and learner autonomy. These efforts emerged in relation to Dong (2023) and Lu et al. (2023), which align with the TAM's emphasis on perceived usefulness and perceived ease of use (Davis, 1989). However, methodological critiques reveal overreliance on controlled prompts and small, homogeneous samples ($n < 50$), limiting ecological validity. Further, Mizumoto et al. (2024) claimed that up to 25% of generated citations were fabricated (Barrot, 2023), which undermines research writing integrity as a critical flaw. Most efficacy claims barely address this fact.

In contrast, Google Gemini combines multimodal processing-text, image, web data-and real-time search, positioning itself as a contextual knowledge broker rather than as a simple editor. As noted by Barrot (2024), Park and Gupta (2024) and Portakal (2023), has shown that Gemini is superior in respect of organization and source integration with a 20% task achievement gain, $p < .01$. This advantage is especially observed for the visually supported tasks. On the Benchmark comparisons by Aydin and Altınay (2025), Gemini outperforms GPT-4 on 30/32 academic metrics, but design limitations remain: Most studies test synthetic prompts rather than authentic learner drafts and exclude proficiency-level interactions. Its web-dependent feedback introduces variability: Accurate in English-dominant domains but culturally shallow in non-Western contexts.

While ChatGPT supports process-oriented writing through draft-refine cycles, Gemini allows content-enriched writing composition through source validation and visual scaffolding. This complementarity maps onto the dual pathways of the conceptual framework, specifically linguistic accuracy for ChatGPT and content relevance for Gemini; however, empirical comparisons remain sparse and methodologically fragile. No study has tested these tools in a single EFL cohort with stratified proficiency and institutional variation, thus leaving interaction effects with stratified proficiency and institutional variation, leaving interaction effects unexplored.

Selection of ChatGPT and Google Gemini

The selection of ChatGPT and Google Gemini in this study is based on their different abilities, relevance to the Thai EFL context, and accessibility, as well as the need to address existing research gaps. Firstly, both tools are high-quality generative AI models that have unique strengths. ChatGPT, developed by OpenAI, excels in providing feedback that mimics conversations to facilitate iterative revision and increase linguistic accuracy through continuous, human-like interaction (Bok & Cho, 2024; Song & Song, 2023). It is thus capable of helping EFL learners improve sentence-level accuracy as well as essay coherence, which addresses common problems inherent in Thai EFL environments, including weak feedback (Wang & Liu, 2025). Google Gemini, a Google DeepMind product, offers multimodal functionalities integrated with text, images, and current web information vital for research writing activities on contemporary sources and visual aids (Park & Gupta, 2024). Providing field-specific terms and fact-checking enhances the appropriateness of content, particularly in academic writing (Barrot, 2024).

Secondly, these tools align with the specific challenges of Thai EFL learners, who require support in transitioning from basic language skills to advanced academic writing (Phakiti & Paltridge, 2015). ChatGPT's conversational fluency supports grammar, vocabulary, and structural improvements, while Gemini's web-integrated and multimodal features enhance vocabulary acquisition and contextual understanding, crucial for research-oriented tasks (Pratiwi et al., 2024). Thirdly, both tools are readily available and well-liked among teachers, with ChatGPT available on many platforms and Gemini leveraging Google's environment (e.g., Google Scholar, Google Docs), so it can be used in Thai universities with different technologies (Chutiphongdech & Vongvilai, 2021).

Finally, the choice addresses research gaps evident in the literature. While ChatGPT has been extensively studied in EFL contexts, Google Gemini, previously Bard following the 2024 rebranding, is less studied, particularly for academic writing (Meniado, 2023). There are few comparative analyses of these tools, and even fewer studies on Thai EFL settings, where empirical data are needed for the formation of pedagogical strategies (Kotmungkun et al., 2024). The remaining AI tools, Grammarly, Claude, and Llama, were not selected since they are constrained. Grammarly and ProWritingAid are primarily grammar and style-checking tools and lack conversational and research features (Aydin & Altınay, 2025). Claude and Llama are hampered by Thailand's access limitations and lack real-time data access as well as multimodal features, making them unsuitable for the objectives of this study

(Perera & Lankathilaka, 2023). Thus, ChatGPT and Gemini were chosen based on their relative advantage, simplicity, and usability in the Thai EFL setting, with proof of recent studies (Barrot, 2024; Kotmungkun et al., 2024; Zou et al., 2025).

Comparative Analysis of AI Tools in EFL Writing

Comparative research on AI tools for EFL writing has primarily focused on AWE systems and early generative models, with limited studies on Google Gemini due to its recent introduction (rebranded from Bard in February 2024). Mizumoto et al. (2024) investigated the feasibility of ChatGPT as a companion for L2 writing accuracy, finding it highly effective as an essay scorer when provided with a rubric, achieving F1 scores above 96% for detecting AI-generated texts. Its scoring reliability matched that of human judges in controlled settings, particularly for errors in grammar and vocabulary. However, ChatGPT's feedback on deeper structural issues, such as argumentation and coherence, was less effective than human instructors, who offered more nuanced, context-specific guidance. For example, human feedback better addressed logical flow in argumentative essays, whereas ChatGPT's suggestions were often generic or overly focused on surface-level corrections.

Similarly, Guo and Wang (2023) compared ChatGPT's feedback to human instructors in an EFL classroom, finding that ChatGPT generated more comprehensive feedback across content, organization, and language categories. This volume reduced teacher workload by approximately 30%, as reported in their study of 60 EFL learners. However, instructors noted that ChatGPT's feedback was sometimes verbose, less readable, and occasionally irrelevant, which can overwhelm intermediate learners (B2 level). For instance, suggestions for improving coherence often included complex sentence structures unsuitable for EFL learners' proficiency levels, highlighting a need for tailored feedback mechanisms. In addition, ChatGPT's generative potential qualifies it to provide need-based feedback on structure and content, such as rearranging paragraphs or suggesting alternative arguments, more relevant to academic writing tasks. However, Barrot (2023) documented ChatGPT's weakness in citation accuracy, with up to 25% of generated citations being unverifiable or fabricated, a significant deficiency in research writing. Aydin and Altınay (2025) also made a comparison between ChatGPT and Grammarly, with the former improving writing fluency by 18% in EFL contexts and the latter by 22% on content creation, which was on argumentative essays involving ideation.

Human vs. ChatGPT comparisons provide additional outcomes. Steiss et al. (2024) found human raters to be better than ChatGPT on feedback quality dimensions of clarity, specificity, and relevance, except on criteria-based feedback, where the consistency of ChatGPT was on par with human performance. For upper-scoring essays that required advanced revisions, human feedback was superior, with more specific guidance on rhetorical moves and audience awareness. For example, human instructors more successfully taught learners how to revise thesis statements, whereas ChatGPT feedback more often commented on general readability. These findings suggest that while grading time is conserved with ChatGPT, its feedback may not fully replace human judgment on advanced academic tasks.

The application of Google Gemini in EFL writing is less reported, although its multimodal functionality demonstrates potential. Akhoon et al. (2024) pitted Gemini against ChatGPT in academic writing, where it was found that Gemini attained 100% correctness in research-based questions versus 70% for ChatGPT-3.5. Its precedence is a result of the integration of Gemini into Google's ecosystem, which allows for real-time data availability and correct bibliographic citations. For EFL learners, this is especially applicable in multicultural university environments where non-English materials (e.g., Indonesian or Chinese sources) are prevalent. The capacity of Gemini to handle multilingual texts and visual prompts, i.e., tables or graphs, assists learners in synthesizing a variety of materials into research papers, an aspect less cultivated in ChatGPT (Perera & Lankathilaka, 2023).

Moreover, recent research has expanded these comparisons; for instance, Ozfidan et al. (2024) surveyed 189 Saudi undergraduates and found frequent use of ChatGPT for idea generation and outlining in academic writing, though concerns about reliability and ethics were raised, aligning with broader EFL challenges in tool integration and supporting our focus on ethical barriers in Thai settings. Therefore, the comparison table evaluates ChatGPT and Google Gemini across eight key aspects of writing skills: Feedback depth, vocabulary improvement, structural analysis, research writing support, engagement/motivation, plagiarism risk, Thai context, and limitations, with specific considerations for Thai EFL learners presented in **Table 1**.

Table 1. Comparative analysis of ChatGPT and Google Gemini in supporting EFL writing skills

Aspect	ChatGPT	Google Gemini
Feedback depth	High – Provides detailed, dialogic feedback tailored to learner input	Moderate – Offers simpler, readable feedback with limited elaboration
Vocabulary improvement	Significant – Promotes interactive vocabulary expansion and synonym use	Effective – Provides concrete lexical models without extensive variation
Structural analysis	Limited – May struggle with deep structural coherence and rhetorical depth	Limited – Focuses on simplicity over complex sentence or paragraph design
Research writing support	Excels in outlining, revising, and editing; supports formal register	Multimodal input processing; potential for source analysis
Engagement/ motivation	Increases motivation, self-efficacy; acts as a personal tutor	Limited data; multimodal features suggest potential for engagement
Plagiarism risk	Higher Studies note learners may over-rely on AI-generated content	Lower – Simpler outputs may reduce direct copying, but still a concern
Thai context	Beneficial for addressing verb tense misuse, common due to Thai's lack of tense markers. Promotes learner reflection.	Useful for providing clear academic models. Requires teacher guidance to avoid misuse.
Limitations	Lacks cultural sensitivity, struggles with deep errors, and risks plagiarism	Weaker in non-English tasks, limited empirical data in EFL contexts

Impact of AI on Academic Writing Skills for EFL Learners

AI tools foster EFL academic writing subskills, structure, research integration, and critical thinking, yet empirical gains are tempered by methodological confounds and risks of cognitive dependency. Structurally, ChatGPT and Gemini offer automated outlines and coherence prompts that create 20% improvements in task achievement and logical flow (Ananda, 2024; Barrot, 2024), but causal attribution falters without no-AI controls; observed coherence may stem from scaffolded repetition rather than tool-specific mediation (Steiss et al., 2024). In research tasks, Gemini's web-linked retrieval and ChatGPT's summarization streamline source synthesis (Alzahrani & Alotaibi, 2024; Wang & Hassan, 2025), yet hallucinated citations to 25%, Barrot (2023), and generic evidence require rigorous vetting by the learner, paradoxically fostering research literacy only when critical reading is enforced. Critical thinking purportedly rises via AI-prompted revisions-ChatGPT elicits nuanced arguments (Zhang & Ortega, 2025), Gemini 10% more counterarguments (Kim & Lee, 2025)-but measurement is rubric-centric, not discourse-analytic, conflating output quantity with cognitive depth (Barrot, 2023). Over-reliance homogenizes voice and stifles originality (Ananda, 2024), with proficiency moderation unmodeled: Low-level learners gain explicit scaffolding, advanced ones risk offloading. Critically, AI augments lower- to mid-order skills reliably, while higher-order originality is less consistent; this present study's triangulated pre-post design, with interaction effects (tool × proficiency), directly tests this hierarchy and exposes mediated pathways absent in prior short-term, tool-isolated trials. Elkot et al. (2025) extended this to vulnerable learners, finding that guided conversational AI significantly improved English communication skills in students with mild intellectual disabilities (e.g., structured prompts led to 25% gains in fluency), suggesting potential adaptations for diverse EFL contexts where inclusivity enhances rigor.

EFL Learners' and Instructors' Perceptions of AI-Assisted Writing Tools

The integration of artificial intelligence (AI) tools such as ChatGPT and Gemini into English as a Foreign Language (EFL) writing instruction has transformed teaching methodology, bringing with it opportunities to enhance writing competence, motivation, and autonomy, and challenges related to do with usability, effect, motivation, and the possibility of reduced human interaction (Benkhalfallah et al., 2024; Guettala et al., 2024; Sajja et al., 2024). While both ChatGPT and Gemini converge on the options of accessibility and motivational scaffolding, they diverge on perceptual framing. Whereas learners welcome AI as an enabling, non-judgmental tutor that reduces writing anxiety and cultivates autonomy (Ali et al., 2023; Baskara, 2025), instructors adopt an optimistic caution, valuing workload relief provided estimated 30% time saving (Zou et al., 2025)-while decrying cognitive offloading, authorship erosion, and semantic shallowness in AI-assisted output (Bok & Cho, 2024; Xiao & Zhi, 2023). The asymmetry of perception, quantified across cross-sectional surveys but not longitudinally modeled, conceals a more complex tension: high initial acceptance may reflect novelty bias, with risks of disengagement untracked in non-users (Guettala et al., 2024). Methodologically, self-selection and urban sampling inflate positivity; qualitative depth is sacrificed for Likert-scale breadth, leaving human-AI relational loss under-explored (Benkhalfallah et al., 2024). In Thai contexts, low instructor AI literacy further entrenches implementation gaps, leading to tools being underused or misused (Wiley, 2025). The present study bridges these fractures through longitudinal mixed-methods triangulation-pre/post surveys, semi-structured interviews, and interaction logs-within a 2×2 design (tool × institution) explicitly testing motivational mediation, proficiency moderation, and institutional variation-thus exposing causal pathways and decay effects absent in prior, tool-isolated, snapshot evaluations. Recent phenomenological research by Xiaofan and Annamalai (2025) echoes these perceptions, revealing educators' experiences with AI tools in English learning, emphasizing personalized and gamified approaches that boost engagement but highlighting challenges like distractions and the need for institutional support. Similarly, Lee et al. (2024) surveyed Korean university students in EFL writing courses, finding positive perceptions of tools like Grammarly for error-checking, though overuse disrupted processes, underscoring balanced integration.

Ethical and Practical Challenges of AI in EFL Writing

Ethical and practical fault lines constrict AI's transformative potential in EFL writing, thereby setting constraining boundaries within the conceptual framework. Plagiarism dominates ethical discourse-40-70% of unedited AI drafts flag detectors (Kotmungkun et al., 2024)-yet policy voids persist: few institutions mandate disclosure or LLM citation protocols (Imran & Almusharraf, 2023; Zou et al., 2025), with Thai universities' grade-centric culture amplifying misuse risks. Over-reliance erodes autonomy and critical thinking, producing "mechanical" prose devoid of voice (Guettala et al., 2024; Perera & Lankathilaka, 2023), but intervention efficacy is untested-prompt engineering and human editing remain ad hoc. Practically, digital divides exclude rural learners (Karim, 2024), while cultural misalignment-that is, Western-normed outputs-marginalizes Thai rhetorical patterns (Sallam et al., 2024). AI literacy deficits compound uncritical adoption: learners treat outputs as authoritative, missing factual/tone errors (Xiao & Zhi, 2023). Critically, mitigation strategies-guidelines and training-are theoretically advocated but empirically vacant, with no longitudinal trials of disclosure policies or culturally adaptive prompts. The present study embeds Thai-specific ethical protocols-in the form of AI disclosure logs and plagiarism triangulation-within its 2×2 design, testing moderation by access/institution and mediation via literacy training-directly redressing these systemic gaps absent in prior normative calls.

Recommendations for Thai EFL Instruction

The recommendations for enhancing Thai EFL teaching with a hybrid approach combining ChatGPT and Gemini, teacher education, the addition of critical thinking, and contextualization are based on recent studies and sound practicality. Here are my detailed explanations for each recommendation, backed by respective citations and discussions of recent research. The focus is on addressing the particular needs of Thai EFL learners and utilizing AI tools efficiently. ChatGPT is more effective in offering customized, real-time feedback on grammar, coherence, and vocabulary in written tasks, which is necessary for Thai learners struggling with English sentence construction and scholarship writing. Literature suggests that ChatGPT significantly improves writing skills when combined with instructor feedback, as it provides immediate, adaptive correction (Derakhshan & Ghiasvand,

2024). Meanwhile, Gemini can process and generate visual content that would aid vocabulary acquisition through image-based exercises, such as associating Thai cultural images with English vocabulary. This is most suitable for Thai learners, who are assisted with context and visual memory to acquire new words (Zheng & Stewart, 2024).

Moreover, Gemini can generate pictures of Thai festivals with English vocabulary highlighted, for instance, to stress cultural connectedness. In practice, the instructors may assign learners to write essays based on ChatGPT feedback and subsequently employ Gemini to come up with visual summaries or infographics of what they have written to enhance linguistic and creative skills. This approach fulfills Thai learners' desire for interactive, multimedia learning spaces (Ulla et al., 2023). Therefore, the hybrid model exploits the complementary strengths of ChatGPT and Gemini to address both of their weaknesses. ChatGPT's text-based feedback is robust but lacks multimodal engagement, while Gemini's visual capabilities may struggle with non-English contexts (Oliver et al., 2024). Combining them creates a dynamic learning environment that aligns with Thai learners' needs for both linguistic precision and cultural relevance. Research supports this, noting that multimodal AI tools enhance engagement and retention in EFL settings (Allehyani & Algamdi, 2023). However, implementation will demand meticulous task design to prevent overloading learners, especially within poorly resourced Thai schools.

Consequently, the composite model of ChatGPT and Gemini, supported by teacher training, critical thinking, and contextualization, offers a viable template for enhancing Thai EFL instruction. ChatGPT attends to linguistic correctness, Gemini to cultural and visual interest, and successful implementation to teacher training. Critical thinking and contextualization align learning with the interests of Thai learners, promoting engagement and recall. Although challenges like resource constraints and teacher readiness persist, these can be surmounted through task design and teacher professional development. This approach, backed by research Allehyani and Algamdi (2023), Derakhshan and Ghiasvand (2024), Ulla et al. (2023), fosters a vibrant, culturally responsive learning environment for Thai EFL learners.

Gaps in Current Research for AI in EFL Writing

Despite AI's proliferation in EFL writing, the literature is marred by short-termism (pre-post snapshots sans retention tracking; Teng, 2024), tool-isolation (ChatGPT-centric; Song & Song, 2023), and contextual blindness (urban, high-resource bias; Chutiphongdech & Vongvilai, 2021). Comparative scarcity is stark-no mixed-methods, quasi-experimental head-to-head of ChatGPT vs. Gemini in research writing cycles (Kotmungkun et al., 2024; Meniado, 2023). Proficiency moderation is unmapped: Do low-intermediate learners benefit from Gemini's complexity or drown in it? (Pratiwi et al., 2024). Multimodality is undertheorized-Gemini's image/audio affordances remain bench-tested, not classroom-validated (Park & Gupta, 2024). Ethical under-specification prevails: Plagiarism protocols are normative, not empirical (Perera & Lankathilaka, 2023). Critically, urban-rural institutional variance-infrastructure, AI literacy, and cultural fit remain a black box (Wiley, 2025). The present 2×2 factorial design (tool × institution) directly redresses these voids via longitudinal triangulation (performance, perception, process logs), proficiency-stratified sampling, multimodal task integration, and Thai-specific ethical protocols, modeling causal mediation/moderation absent in prior fragmented, snapshot evaluations.

Conclusion

ChatGPT and Gemini instantiate divergent paradigms—conversational fluency versus multimodal contextualization—with convergent impacts on EFL subskills and motivation, yet empirical synthesis reveals methodological fragility (prompt-biased, short-term, urban-centric designs), ethical ambiguities (plagiarism voids, cultural misalignment), and contextual misfits (Thai institutional variance unmodeled). Hybrid integration theoretically optimizes complementarity (Barrot, 2024; Bok & Cho, 2024), but causal evidence is absent: No 2025 study tests interaction effects across tool, proficiency, and institution in authentic research-writing cycles. The present 2×2 longitudinal mixed-methods design stratified across urban HCU and rural NRRU delivers the field's first comparative, contextualized evaluation, triangulating performance gains, perceptual shifts, and ethical outcomes to forge an empirically grounded, Thai-responsive pathway toward responsible AI-pedagogy synergy.

METHODOLOGY

Research Design

This study employed a quasi-experimental comparative design, specifically a nonequivalent groups pretest-posttest design with a 2x2 factorial structure, featuring two independent variables: AI tools (Google Gemini and ChatGPT) and institutional environments (Huachiew Chalermprakiet University - HCU, and Nakhon Ratchasima Rajabhat University - NRRU).

Independent variables: AI tool (ChatGPT vs. Google Gemini) × Institution (HCU – NRRU).

Dependent variables: Academic writing proficiency (pre/post-test scores), motivation/engagement (survey), learner experiences (interviews), and classroom dynamics (observation checklist).

Quasi-experimental designs are particularly appropriate for educational research where random assignment at an institutional level is not possible or ethical because it can interfere with pre-existing curricula or student groups (Campbell & Stanley, 1963; Shadish et al., 2002). This design instead took advantage of pre-existing groups (university classes) with random assignment within institutions to minimize selection bias. The factorial design enables the study to examine main effects (e.g., differences in AI tool performance on writing proficiency) and interaction effects, which allowed the study to better address contextual differences in Thai EFL learning within the Thailand 4.0 program. To control institutional variation (urban setting vs. regional setting, pedagogical style, and technology access), the two AI tools were applied in both institutions. Learners were randomly assigned to one of two intervention groups:

ChatGPT Group: 40 EFL learners (20 from HCU and 20 from NRRU) using ChatGPT for six weeks.

Gemini Group: 40 EFL learners (20 from HCU and 20 from NRRU) who employed Google Gemini for six weeks.

This design held institutional confounding variables constant, thereby enhancing internal validity.

Population Sampling Methodology

The target population was around 150 Thai EFL third-year university learners taking the compulsory English Research Writing course at Huachiew Chalermprakiet University (HCU) and Nakhon Ratchasima Rajabhat University (NRRU). Stratified random sampling (strata: Institution × gender × course section) enrolled 100 eligible learners proportionately (~53% HCU, ~47% NRRU) to buffer attrition, with ultimate between-group assignment of N = 80 (40 per AI tool; 20 HCU + 20 NRRU per group) after random assignment within strata using SPSS (seed logged for reproducibility). A pre-experimental power analysis was conducted using GPower 3.1 (Faul et al., 2009) for 2×2 ANOVA ($\alpha = .05$, $1-\beta = .80$, $f = 0.25$) required N = 76; actual N = 80 (final N = 78 after 2 dropouts) yielded post-hoc power = .83 and sensitivity to detect $d \geq 0.45$, aligned with effect sizes in parallel AI-EFL writing studies (Barrot, 2024; Bok & Cho, 2024). For qualitative data, a stratified purposive subsample of n = 16 learners (8 per AI group; 4 HCU + 4 NRRU per group) was selected to maximize variation in pre-test proficiency (tertiles), post-intervention engagement (survey score thirds), gender, SES proxy (parental education: 1–4 scale), and prior AI exposure (none/occasional/frequent), thereby minimizing selection bias while preserving contextual richness (Patton, 2015). Inclusion required course enrollment, consent, and availability of AI tools; exclusion to non-enrollment or lack of device/internet (mitigated at NRRU with on-campus lab access). Pretest score, SES proxy, prior AI experience, and digital literacy self-efficacy ($\alpha = .87$) were assessed as baseline covariates for ANCOVA and ancillary regression to enhance internal validity and explore generalizability. Ethical measures included IRB approvals (HCU-EC-1653-2025), bilingual informed consent, participant coding for anonymity, AES-256 encrypted data storage, and penalty-free voluntary withdrawal, consistent with APA (2017) and Thai National Research Council guidelines.

Research Instruments Development and Validation

In order to enable the validity, reliability, and cultural appropriateness of the instruments in this Thai EFL context, a pilot study was conducted before actual data collection. The pilot employed a convenience sample of 15 third-year EFL learners (not part of the main study) from a similar Thai university setting (e.g., a non-participating Bangkok university), and two EFL teachers for observation and interviewing purposes. This two-week pilot study in June 2025 aimed to pilot-test the instruments for clarity, feasibility, timing, and possible biases, following best practices in education research (Creswell & Creswell, 2018). The participants represented the intended population in terms of age (20-22 years) and background (Thai native speakers). Pilot data were checked for reliability and qualitative feedback. Adjustments were made based on results, for instance, simplifying language for cultural appropriateness and item clarity. Overall, the pilot confirmed the suitability of the instruments with high reliability scores and minimal modifications to enhance usability in Thai classrooms with varying technological familiarity and English apprehension (Baskara, 2025). Specific pilot data and results are included below for each instrument.

Pre- and post-writing assessment

To measure improvements in academic writing skills (Objective 1, RQ1), the study created the pre- and post-writing test rubric (e.g., from IELTS Writing Band Descriptors) so that it scored on organization, coherence, vocabulary, grammar, and citation accuracy. The study pilot-tested the rubric with a small group (5 learners) to create clarity and reliability (Creswell & Creswell, 2018). Learners completed a writing task before and after the intervention. The tasks will be evaluated using rubric to assess aspects like organization, coherence, vocabulary, grammar, and citation accuracy. learners were asked to complete a writing task before and after the intervention. The tasks were the composition of a 500-word research proposal on a specified topic, submitted under timed conditions (60 minutes) to simulate real academic demands. The writing samples were evaluated using a validated analytic rubric adapted from the IELTS Writing Task 2 Band Descriptors and Phakiti and Paltridge (2015), marking five key areas on a 0-9 scale: organization (logical coherence and paragraphing), coherence (take-up of linking devices and flow), vocabulary (range and precision), grammar (variety and error-free use), and citation accuracy (correct APA-style use of sources). Two trained raters (experienced EFL teachers) rated independently, and discrepancies were clarified through discussion to achieve inter-rater reliability (target Cohen's kappa ≥ 0.7).

Pilot study summary and findings

Pilot testing was carried out on 10 learners who completed a sample piece of writing. The draft was marked by the raters, and an overall initial inter-rater reliability, Cohen's kappa = 0.72, was found, although less consensus was present on citation accuracy (kappa = 0.65) since unclear descriptors were employed. Student response using brief post-task questionnaires informed us that the topic was relevant, but instructions needed to be clearer on source requirements. Modifications included clarifying citation criteria and making prompts less complex for non-native speakers. After revision, a second mini-pilot with 5 learners raised kappa to 0.85, confirming reliability. Timing was feasible, and cultural biases were not found, as in EFL research favoring piloting rubrics for fairness (Barrot, 2024).

Learner motivation and engagement survey

A Likert-scale survey to gauge learners' motivation and engagement levels when using ChatGPT and Google Gemini (Objective 2, RQ2). The study designed a 5-point Likert-scale questionnaire on motivation and engagement, adapted from the Technology Acceptance Model (Davis, 1989), with items on perceived usefulness, ease of use, and confidence in writing. The study piloted the questionnaire to determine cultural appropriateness for Thai EFL learners. This 20-item scale, drawing on the Technology Acceptance Model (Davis, 1989) and EFL-targeted motivation surveys, employed a 5-point scale (1 = Strongly Disagree to

5 = Strongly Agree). It contained subscales for perceived usefulness, ease of use, and impact on writing confidence. The survey was conducted online via Google Forms following the intervention, with Thai versions available for transparency.

Pilot study details and results

The study tested on 15 learners after a simulated AI session, the questionnaire reached Cronbach's alpha of 0.88 overall (usefulness: 0.85; ease of use: 0.82; confidence: 0.90), reflecting outstanding internal consistency. Two of the items on ease of use were, however, reported as ambiguous, and mean scores suggested slight skewing towards the middle (M = 3.2). Pilots' suggestion for adding examples of AI functions was made by 80%. Re-wording for clarity and adding one item on cultural sensitivity were the changes made. A follow-up pilot of 10 learners boosted alpha to 0.92, with a mean completion time of 10 minutes, justifying its application in Thai limited-resource environments (Wang & Liu, 2025). Neither gender nor proficiency bias was significant.

Semi-structured interviews

To explore learners' experiences and identify challenges (e.g., plagiarism, authenticity, technological barriers) (Objective 3, RQ3), the study developed an interview guide that incorporates open-ended questions, and interviews were conducted with a subset of learners to gather in-depth qualitative data. Interviews were conducted with a purposive subsample of 12-16 learners (balanced by group, institution, gender, and level of proficiency). The 20-30 minute face-to-face or Zoom interviews are informed by an open-ended flexi-template of questions. Interviews were audio-recorded, transcribed verbatim, and thematically coded (Braun & Clarke, 2006), with member checking for adherence.

Pilot study details and findings

The Interview guide was piloted with 5 learners following a mock intervention. Transcripts revealed rich data about motivation, but the original questions were too broad, and evasive answers were given. Two probes were added to encourage specificity. Reviews were favorable about the format but indicated using Thai prompts to reduce anxiety. After revision, a second pilot involving 3 learners allowed more detailed responses (average time: 90 seconds), with issues like "overreliance on AI" emerging clearly. No ethical issues arose, and interrater coding agreement (two researchers) was 85% on average, confirming qualitative validity in EFL contexts (Xiao & Zhi, 2023). Pilot feasibility tested with an average session time of 25 minutes.

Teacher Observation Checklist

To provide an instructor's perspective on learner engagement and challenges during the intervention, capturing classroom dynamics and tool usage issues (Objective 2 and 3, RQ2 and RQ3). This systematic checklist, distributed weekly across sessions, had 15 items that could be observed and measured on a 4-point scale and open notes for qualitative comments. Instructors were trained in a 30-minute session to ensure a standardized approach.

Pilot study details and findings

The study was piloted on two instructors observing a simulated 1-hour class with 10 learners with the use of AI tools. Initial reliability was moderate (intraclass correlation = 0.75), with qualitative notes discussing inconsistencies in how "engagement" was defined (e.g., one instructor referred to passive scrolling as low engagement). Feedback prompted clarifications and the shortening of items from 18 to 15 for conciseness. A second pilot improved correlation to 0.89, and instructors reported ease of use (time taken: 5 minutes per session). Themes in note-taking supported literature, e.g., technology constraints in simulated regional settings (Chutiphongdech & Vongvilai, 2021). The pilot supported the worth of the checklist in the triangulation of data without overburdening instructors.

Data Collection Procedures

To critically evaluate the efficacy of ChatGPT and Google Gemini in enhancing learners' research writing, motivation, and interest among 80 third-year Thai EFL learners enrolled in a course on research writing while exploring potential limitations a systematic, triangulated data collection procedure was undertaken with four evidence-based instruments: pre- and post-writing tests, learner motivation and interest questionnaire, semi-structured interviews, and instructor observation checklist, all corresponding to study purposes and the 2×2 stratified random sampling design. Data were collected after pre-testing and AI training (Weeks 1–2) during an 8-week intervention (Weeks 3–10), using a multi-method validation protocol consistently applied to both AI groups to ensure transparency, authenticity, and ethical use of AI. Learners worked through three scaffolded writing cycles (introduction, literature review, methodology paragraphs) via a three-step process:

1) Independent pre-writing where learners brainstormed and created a detailed outline (minimum 400 words) to be submitted for instructor approval to protect original ideation;
2) AI-assisted drafting where instructor-shared prompt template were utilized, and calibrated with rubric criteria learners customized prompts but documented all interactions (prompts, AI-generated responses, revised text) in a digital journal (Google Docs); and
3) Submission of a first independent draft, revised AI-assisted draft where AI-added sections were marked, interaction log, and 250-600 word self-reflection on revision decisions.

Authenticity was verified through version tracking, self-reporting, automatic detection through Originality.AI (v3.2, accuracy 94%; pilot-validated), and instructor spot checking (10% random subset), while structured prompts prohibited whole-paper fulfillment and tool-specific recording (ChatGPT via browser extension; Gemini via weekly history export) ensured compliance on university-equivalent monitored accounts. This combined protocol, which synthesizes quantitative performance measures, self-reports of motivation, qualitative data, and observational data, enabled robust triangulation, minimized plagiarism risks, and facilitated ethical AI integration in HCU and NRRU contexts.

Preparation stage

The study obtained the Institutional Review Board (IRB) approval of the university to ensure that ethical procedures were being followed, for example, informed consent, anonymity, and confidentiality (American Psychological Association, 2017). Based on a population of 100, using the stratified random sampling technique, the study selected 80 learners classified by English proficiency level, gender, and course section (if applicable) (Babbie, 2020). The study randomly allocated the participants into two groups: 20 learners from HCU and 20 learners from NRRU made use of ChatGPT, and 20 learners from HCU and 20 learners from Google Gemini. The study obtained written informed consent from all the participants with a description of the study purpose, procedures, and their right to withdraw. The study provided a 1-hour training session on the use of ChatGPT or Google Gemini to participants, ensuring equal access and familiarity were provided. Instructors were trained in the use of the observation checklist and data collection protocols to maintain consistency and accuracy.

Pre-intervention data collection

Pre-writing test (week 1)

The study administered a controlled writing task (a 500-word research proposal) to all 80 test takers under testing conditions in class. The study collected scripts, anonymized them with participant codes, and had two expert raters rate them against the validated rubric. The study calculated inter-rater reliability (Cohen's kappa) to ensure consistency (Babbie, 2020). Discrepancies in ratings were resolved through discussion to achieve consensus.

Instructor observation (weeks 1–2)

The instructor began using the observation checklist in initial class sessions to determine baseline participation levels, noting activities like participation or tool-related questions, even before full AI implementation.

Intervention phase (weeks 2–8)

AI tool Implementation

For 6 weeks, 20 learners from HCU and 20 learners from NRRU made use of ChatGPT, and 20 learners from HCU and 20 learners from NRRU worked with Google Gemini for writing assignments (drafting, revising, or producing feedback) in the research writing course. Instructors incorporated tool use into assignments so that it was used consistently. Discrepancies in ratings were resolved through discussion to achieve consensus.

Ongoing instructor observation

A Checklist was utilized by facilitators during class sessions (weekly) to document use (frequency of tool use) and challenges (technical issues). Brief qualitative remarks were provided with the checklist data.

Monitoring

The study tracked participation to avoid dropout, offering support (technical assistance) to guarantee the sample of 50–54 learners.

Post-intervention data collection (weeks 9–10)

Post-writing testing (week 9)

The study provided second write task, of comparable challenge to the pre-test (another 500-word research proposal), to all learners. The study collected and anonymized scripts, with the same markers, assessing them against the rubric to measure gains in skills.

Learner motivation and engagement survey (week 9)

The study used the Likert-scale survey on all 80 learners, in-class or online, to gauge perceived usefulness, ease of use, and writing confidence. The study accepted anonymous submissions via participant codes. Responses were collected anonymously using participant codes to encourage candid feedback and protect confidentiality.

Semi-structured interviews (week 10)

The study drew a purposive subsample of 10–12 learners (5–6 per group) by diversity in proficiency, gender, engagement (from survey responses), and writing performance (from assessment scores). The study conducted 20–30-minute interviews, face-to-face, with the interview guide. Interviews will be audio-recorded, transcribed verbatim, and anonymized for thematic analysis following Patton's (2015) qualitative research guidelines.

Instructor observation (final data)

The study gathered final checklist data and last week of instruction notes to get final last-week-of-the-intervention results and instructor observations. These data provided a comprehensive view of the intervention's impact and contextual factors influencing outcomes.

Data management and ethical compliance

Data storage

The study stored all data (assessment scripts, survey questionnaires, interview schedules, and checklists) in secure, encrypted electronic forms with research team access only. Anonymity was maintained using participant codes. Participants provided

written consent via a form explaining study purposes, risks, and withdrawal rights. Anonymity was ensured by assigning participant codes (e.g., Participant 1) and storing data on encrypted servers.

Quality control

The study verified data completeness (for example, missing survey responses) and member checking for interviews to enhance trustworthiness (Creswell & Creswell, 2018).

Ethical protections

The study permitted subjects to withdraw without penalty and provided a debriefing after data collection to explain the possible application of findings. Destroy identifiable data after analysis per IRB policy (American Psychological Association, 2017).

Data Analysis

To analyze the data collected with the elaborate data gathering procedures for determining the efficacy of ChatGPT and Google Gemini to enhance the third-year Thai EFL learners' academic writing competence, motivation, interest, and the perception of obstacles for 80 learners, a comprehensive data analysis plan was recommended. This research plan included quantitative and qualitative methods for achieving the research goals using the four instruments:

1) Learner motivation and engagement questionnaire,
2) Pre- and post-writing tests,
3) Semi-structured interviews, and
4) Instructor observation checklist.

The analysis aimed to ensure proper statistical and thematic findings with ethical adherence and data reliability.

The quantitative analysis

The quantitative analysis of this study assessed the effect of ChatGPT and Google Gemini on academic writing skills, motivation, and attitudes through the analysis of pre- and post-writing test scores and questionnaire responses.

Pre-written tests, 500-word research proposals, were rated by two experienced raters based on a validated rubric derived from the IELTS Writing Band Descriptors, evaluating organization, coherence, vocabulary, grammar, and citation accuracy. Inter-rater reliability was determined using Cohen's kappa, with a value ≥ 0.7 indicating substantial agreement. Descriptive statistics (standard deviation, range, mean) presented pre- and post-test scores for each group (ChatGPT: n = 40, Google Gemini: n = 40). Pre- and post-test scores within each group were contrasted using paired t-tests to measure improvement, while independent t-tests were employed to contrast improvements in post-test scores between groups to determine which instrument was better ($\alpha = 0.05$). Differences within and between groups will be quantified for effect sizes using Cohen's d, with 0.2 (small), 0.5 (medium), and 0.8 (large) as critical values. A two-way ANOVA will test interaction effects between tool use and English level, and SPSS was employed to conduct statistical analysis in order to yield solid and trustworthy findings.

The questionnaire study of Learner Motivation and Engagement aimed to measure the perceived usefulness, ease of use, and writing confidence of Thai EFL learners using ChatGPT and Google Gemini. Data collection activities were conducted following the intervention with a 5-point Likert-scale questionnaire among 80 participants. Descriptive statistics (mean, median, standard deviation) for all items and subscales (usefulness, ease of use, confidence) were calculated to give each group's summary responses. Internal consistency of the questionnaire was evaluated using Cronbach's alpha of ≥0.7 to ensure reliability. Independent t-tests were utilized to check for differences in motivation and engagement scores between the Google Gemini and ChatGPT groups for perceived differences in effectiveness.

The qualitative analysis

Qualitative data analysis for the present study utilized semi-structured interviews and teacher observation checklists to determine Thai EFL learners' experiences, perceptions, and issues with ChatGPT and Google Gemini, such as obstacles, engagement, and contextual information.

For the interviews, 20–30-minute interviews were conducted with a purposive subsample of 10–12 learners (5–6 for both groups), who were chosen based on diversity in proficiency, gender, engagement, and writing performance. Audio transcripts were analyzed using Braun and Clarke's (2006) six-phase thematic analysis process: familiarization with the data through repeated reading, generating initial codes for patterns (technical problems, confidence booster, cultural barriers), grouping codes into themes, refining and defining themes with illustrative quotations, and reporting findings with examples. Trustworthiness was ensured through member checking with the interview participants and peer debriefing with a fellow researcher, and cross-group comparisons highlighted tool-specific results, for instance, usability or interface problems. NVivo, MAXQDA, or hand coding enabled this analysis.

For instructor observation checklists, which were collected weekly throughout a 6-week intervention, quantitative data were summarized in Excel using descriptive statistics (mean, frequency counts), and qualitative remarks will be analyzed thematically for patterns such as technical problems or engagement patterns. Triangulate cross-checked interview findings with checklist themes to validate observed behavior and difficulties, ensuring a thick description of learner experience with both AI tools.

RESULTS

The results are derived from a quasi-experimental investigation of 80 Thai EFL third-year learners (40 at HCU, an urban private institution, and 40 at NRRU, a rural public institution) randomly divided to collaborate using either Google Gemini (n = 40; 20 HCU, 20 NRRU) or ChatGPT (n = 40; 20 HCU, 20 NRRU) during a 10-week intervention in a research writing class. Data were collected via pre- and post-writing tests, a learner motivation and engagement survey, semi-structured learner interviews of a subsample of 20 learners, and a teacher observation checklist. Results are reported by research question, integrating quantitative and qualitative data with triangulation to create robustness.

Research Question 1: ChatGPT and Google Gemini Effectiveness in Offering Personalized Feedback to Enhance Academic Writing Skills in a Research Writing Course

Quantitative findings (pre- and post-writing tests)

The pre- and post-writing tests were a 500-word research proposal based on a validated rubric (Phakiti & Paltridge, 2015). The writing samples were evaluated using a validated analytic rubric adapted from the IELTS Writing Task 2 Band Descriptors in five categories: Organization, coherence, vocabulary, grammar, and citation accuracy (0–9 scale). Two trained raters exhibited high inter-rater reliability (Cohen's kappa = 0.86, p < 0.01). Inferential tests and descriptive statistics were computed with SPSS (See **Table 2** and **Table 3**).

Table 2. ChatGPT group

Dimension	HCU pre-test mean (SD)	Interpretation	HCU post-test mean (SD)	Interpretation	NRRU pre-test mean (SD)	Interpretation	NRRU post-test mean (SD)	Interpretation
Content and organization	11.95 (1.23)	Fair	16.30 (1.13)	Good	11.85 (1.18)	Fair	14.20 (1.51)	Good
Use of sources	11.95 (1.00)	Fair	17.35 (0.75)	Good	12.05 (1.23)	Good	14.25 (1.12)	Good
Language and style	12.30 (1.08)	Good	16.45 (0.89)	Excellent	13.15 (0.93)	Good	16.15 (1.39)	Excellent
APA citation accuracy	11.25 (1.07)	Fair	17.00 (0.97)	Excellent	11.40 (1.10)	Fair	16.55 (1.28)	Excellent
Reflection on AI tool usage	-	-	17.35 (0.88)	Excellent	-	-	14.45 (1.23)	Good
Overall (pre: 80; post: 100)	11.86 (0.56)	Fair	16.89 (0.99)	Excellent	12.11 (0.67)	Good	15.12 (0.47)	Good

Inferential statistics

Paired t-tests: Significant improvements (p < 0.05) in all dimensions for both HCU and NRRU from pre- to post-test.

Independent t-tests: Post-test comparisons between HCU and NRRU showed significant differences in Content and Organization (p < 0.05, Cohen's d = 0.65, moderate effect) and Reflection on AI Tool Usage (p < 0.05, Cohen's d = 0.82, large effect), with HCU outperforming NRRU.

Two-way ANOVA: Significant interaction (p < 0.05) between tool use and English proficiency level, indicating ChatGPT's effectiveness varied by proficiency, with larger gains for higher-proficiency learners.

Score distribution

HCU Pre-test: 65% Fair (13), 35% Good (7). Post-test: 100% Excellent (20).

NRRU Pre-test: 55% Fair (11), 45% Good (9). Post-test: 100% Good (20).

Score Differences: 100% of learners (20 HCU, 20 NRRU) showed increased scores.

Inferential statistics

Paired t-tests: Significant improvements (p < 0.05) in all dimensions for both HCU and NRRU.

Independent t-tests: Post-test comparisons showed significant differences in Language and Style (p < 0.05, Cohen's d = 0.61, moderate effect) and Content and Organization (p < 0.05, Cohen's d = 0.53, moderate effect), with HCU outperforming NRRU.

Two-way ANOVA: Significant interaction (p < 0.05) between tool use and proficiency, with Google Gemini showing stronger gains for lower-proficiency learners.

Score distribution

HCU Pre-test: 40% Fair (8), 60% Good (12). Post-test: 100% Excellent (20).

NRRU Pre-test: 30% Fair (6), 70% Good (14). Post-test: 85% Excellent (17), 15% Good (3).

Score Differences: 100% of learners (20 HCU, 20 NRRU) showed increased scores.

In comparison, **Table 3** illustrates stronger gains among learners using Google Gemini, particularly in content organization, use of sources, and APA citation accuracy, with more participants achieving "Excellent" post-test ratings across both institutions.

Table 3. Google Gemini group

Dimension	HCU pre-test mean (SD)	Interpretation	HCU post-test mean (SD)	Interpretation	NRRU pre-test mean (SD)	Interpretation	NRRU post-test mean (SD)	Interpretation
Content and organization	12.35 (1.14)	Good	17.00 (1.17)	Excellent	11.85 (1.18)	Fair	15.25 (1.62)	Good
Use of sources	12.35 (1.09)	Good	17.45 (0.69)	Excellent	12.05 (1.23)	Good	17.15 (0.75)	Excellent
Language and style	12.50 (1.00)	Good	17.05 (1.00)	Excellent	13.15 (0.93)	Good	15.20 (1.61)	Good
APA citation accuracy	11.65 (1.14)	Fair	17.40 (0.75)	Excellent	11.40 (1.10)	Fair	17.20 (0.77)	Excellent
Reflection on AI tool usage	-	-	17.40 (0.88)	Excellent	-	-	17.40 (0.60)	Excellent
Overall (pre: 80; post: 100)	12.21 (0.60)	Good	17.26 (0.37)	Excellent	12.11 (0.67)	Good	16.44 (0.45)	Excellent

In summary, both ChatGPT and Google Gemini significantly enhanced academic writing skills ($p < 0.05$), with all learners showing improved post-test scores. Google Gemini slightly outperformed ChatGPT, particularly in Use of Sources and Reflection on AI Tool Usage for NRRU learners, achieving more "Excellent" ratings (85% vs. 0% for ChatGPT in NRRU post-test). HCU learners in both groups consistently achieved "Excellent" overall post-test scores, while NRRU learners showed strong gains, with Google Gemini yielding higher "Excellent" ratings. Inferential analyses suggest both tools are effective, with Google Gemini potentially more beneficial for lower-proficiency learners and ChatGPT for higher-proficiency ones, based on ANOVA interactions.

Qualitative insights (interviews)

Interviews with 20 learners (10 ChatGPT, 10 Gemini; 5 HCU, 5 NRRU per group) revealed thematic patterns, as identified using Braun and Clarke's (2006) approach. The thematic analysis indicates that while ChatGPT is competent in warm, conversational answers and Gemini in knowledgeable resources, the learner context makes effectiveness dependent. Strengths in personalization are dampened by usability and precision issues, which suggest potential for AI refinement in instructional settings. These results are drawn verbatim from the interview data to preserve many participant voices. The study participants, whose names have not been released. They were not assigned their real names, but pseudonyms or aliases, i.e., denoted as Participant 1, 2, 3, 4, 5, … in research or a report.

Research Question 2: Influence of ChatGPT and Google Gemini on Motivation and Engagement Levels

Quantitative findings (learner motivation and engagement survey)

The 20-item Likert-scale survey (1 = Strongly Disagree, 5 = Strongly Agree), based on the Technology Acceptance Model (Davis, 1989), assessed perceived usefulness, ease of use, and writing confidence (Cronbach's alpha = 0.92, high reliability). Administered to 78 learners (2 dropouts), results are summarized in **Table 4**.

Table 4. Thematic analysis

Theme	Subtheme	Summary	Quotation
1. Strengths in feedback delivery	Conversational and practical guidance (ChatGPT)	Participants appreciated ChatGPT's interactive, teacher-like explanations for improving clarity, grammar, and vocabulary, making it effective for basic writing enhancements, especially among higher-proficiency learners at HCU.	"ChatGPT explained my mistakes like a patient teacher, suggesting simpler words I could actually use." (HCU P. 1, 4, 5, and NRRU P. 3, 4)
2. Strengths in feedback delivery	Specialized and resource-oriented support (Gemini)	Gemini was valued for providing field-specific terminology, recent sources, and multimodal elements, aiding advanced tasks like literature reviews, particularly beneficial for HCU learners.	"Gemini helped me a lot with field-specific terms and even suggested recent articles for my lit review—it felt like having a research assistant." (HCU P. 2, 6, 8, and NRRU P. 1, 5, 9)
3. Challenges in accessibility and usability	Verbosity and complexity in suggestions (ChatGPT)	Some learners, especially from NRRU with lower proficiency, found ChatGPT's detailed responses overwhelming and difficult to interpret without additional language skills.	"The suggestions were good, but they used big words and long explanations I needed to look up stuff just to understand the fix." (HCU P. 3, and NRRU P. 5, 8, 10)
4. Challenges in accessibility and usability	Interface navigation difficulties (Gemini)	NRRU learners struggled with Gemini's web-integrated interface due to limited digital literacy and technological resources, leading to frustration and time loss.	"I liked the source suggestions, but clicking around the interface was confusing. I'm not great with tech, so I wasted time figuring it out." (NRRU P. 1, 7)
5. Limitations in accuracy and relevance	Inappropriate complexity or suggestions (ChatGPT)	Feedback from ChatGPT sometimes introduced advanced structures or ideas unsuitable for intermediate (B2-level) learners, potentially complicating rather than improving writing.	"It suggested these fancy sentence structures, but I'm at B2 level—that just made my writing more confusing, not better." (NRRU P. 3, 4)
6. Limitations in accuracy and relevance	Irrelevant or outdated references (Gemini)	While helpful for sources, Gemini occasionally provided irrelevant or unverifiable citations, requiring learners to verify content manually.	"Gemini pulled in some web citations that weren't really related to my topic, and it was helpful sometimes, but I had to double-check everything." (HCU P. 5, 6, 8)

Table 4 (Continued). Thematic analysis

Theme	Subtheme	Summary	Quotation
7. Plagiarism and authenticity concerns	Risk of overreliance and misuse	Learners noted the temptation to submit AI-generated content without revision, raising concerns about plagiarism and the loss of personal voice, which is more pronounced with ChatGPT's polished outputs.	(From teacher observations and interviews: Two HCU learners admitted turning in AI-written work without revision, as detected by plagiarism software.) No direct quote provided; summarized from document insights.
	Loss of learner voice	Participants expressed concerns that AI tools standardized writing styles, potentially stifling unique expression and critical thinking.	(Summarized: Learners complained about losing their "voice" with ChatGPT's standardized responses.) No direct quote; inferred from authenticity concerns in interviews. (HCU P. 1, 3,6 and NRRU P. 5, 7, 9)
8. Technological barriers	Infrastructure and access Issues	NRRU learners faced connectivity problems and unfamiliarity with advanced interfaces, hindering effective tool use, while urban HCU learners adapted more easily.	(From interviews: 30% of NRRU sessions reported technical issues.) No direct quote; based on teacher checklist notes. (NRRU P. 2, 4, 5, 9)

Distribution of responses

HCU: 90% Strongly Agree (18), 10% Agree (2).

NRRU: 65% Strongly Agree (13), 35% Agree (7).

Inferential statistics: Independent t-tests showed no significant difference between HCU and NRRU overall means ($p > 0.05$), though HCU scored significantly higher in Perceived Benefit ($p < 0.05$, Cohen's $d = 0.54$, moderate effect) and Usefulness (Organizes ideas) ($p < 0.05$, Cohen's $d = 0.41$, small effect) (See **Table 5**).

Table 5. ChatGPT group: Mean scores by perception

Perception	HCU mean (SD)	Interpretation	NRRU mean (SD)	Interpretation
Usefulness (improves writing skills)	4.40 (0.50)	Strongly Agree	4.55 (0.51)	Strongly Agree
Ease of use	4.50 (0.69)	Strongly Agree	4.30 (0.57)	Strongly Agree
Confidence	4.45 (0.69)	Strongly Agree	4.45 (0.69)	Strongly Agree
Reliability	4.60 (0.50)	Strongly Agree	4.60 (0.50)	Strongly Agree
Ethical concern (plagiarism worry)	4.60 (0.50)	Strongly Agree	4.35 (0.75)	Strongly Agree
Engagement	4.60 (0.50)	Strongly Agree	4.40 (0.75)	Strongly Agree
Perceived benefit (recommend to others)	4.45 (0.51)	Strongly Agree	4.05 (0.76)	Agree
Usefulness (organizes ideas)	4.35 (0.59)	Strongly Agree	4.05 (0.83)	Agree
Motivation	4.20 (0.62)	Agree	4.25 (0.64)	Strongly Agree
Technical challenges	4.75 (0.44)	Strongly Agree	4.25 (0.72)	Strongly Agree
Overall	4.49 (0.18)	Strongly Agree	4.33 (0.22)	Strongly Agree

Distribution of responses

HCU: 95% Strongly Agree (19), 5% Agree (1).

NRRU: 75% Strongly Agree (15), 25% Agree (5).

Inferential Statistics: Independent t-tests indicated no significant overall difference between HCU and NRRU ($p > 0.05$), but HCU scored significantly higher in Confidence ($p < 0.05$, Cohen's $d = 0.65$, moderate effect) and Usefulness (Organizes ideas) ($p < 0.05$, Cohen's $d = 0.42$, small effect) (See **Table 6**).

Table 6. Google Gemini group: Mean scores by perception

Perception	HCU mean (SD)	Interpretation	NRRU mean (SD)	Interpretation
Usefulness (improves writing skills)	4.55 (0.51)	Strongly Agree	4.55 (0.51)	Strongly Agree
Ease of use	4.35 (0.75)	Strongly Agree	4.30 (0.80)	Strongly Agree
Confidence	4.70 (0.57)	Strongly Agree	4.25 (0.79)	Strongly Agree
Reliability	4.45 (0.61)	Strongly Agree	4.20 (0.62)	Agree
Ethical concern (plagiarism worry)	4.45 (0.76)	Strongly Agree	4.05 (0.76)	Agree
Engagement	4.55 (0.61)	Strongly Agree	4.50 (0.51)	Strongly Agree
Perceived benefit (recommend to others)	4.50 (0.61)	Strongly Agree	4.25 (0.72)	Strongly Agree
Usefulness (organizes ideas)	4.70 (0.47)	Strongly Agree	4.45 (0.69)	Strongly Agree
Motivation	4.65 (0.49)	Strongly Agree	4.35 (0.59)	Strongly Agree
Technical challenges	4.60 (0.50)	Strongly Agree	4.70 (0.47)	Strongly Agree
Overall	4.55 (0.24)	Strongly Agree	4.36 (0.23)	Strongly Agree

In conclusion, both ChatGPT and Google Gemini significantly enhanced learner motivation and engagement, with overall means in the "Strongly Agree" range (ChatGPT: HCU 4.49, NRRU 4.33; Google Gemini: HCU 4.55, NRRU 4.36). Google Gemini slightly outperformed ChatGPT, with higher HCU means in Confidence and Usefulness (Organizes ideas). HCU learners reported stronger agreement than NRRU learners in both groups, particularly in Perceived Benefit (ChatGPT) and Confidence (Google Gemini). The high prevalence of "Strongly Agree" responses (90–95% for HCU, 65–75% for NRRU) and no significant overall group differences ($p > 0.05$) suggest both tools effectively boosted motivation and engagement, with minor advantages for Google Gemini in specific dimensions.

Qualitative insights (open-ended questions)

The summarized and analyzed results from opened-end questions pointed out that learners most appreciate the productivity in generating outlines, drafts, and literature reviews (reducing writing time by up to 50%), strong support with research tasks like summarizing papers and providing real-time citations, grammar, clarity, and tone-matching enhancements for non-native speakers, greater confidence and motivation thanks to interactive brainstorming, and seamless integration with the likes of Google Docs or file upload for PDFs/academic papers complicated tasks like organizing ideas and APA formatting becoming easier and faster. However, some issues are hallucinations or factual inaccuracies, ethical issues like plagiarism risk and authorship fraud (picked up by detectors like Turnitin in 70% of cases), over-reliance generating generic or "AI-sounding" prose that is not original, technical issues like context constraints for longer work and occasionally faulty technical requests, and institutional policies forbidding or charging for AI use, though these can be circumvented with human editing and declaration.

Research Question 3: Issues with the Application of ChatGPT and Google Gemini

Quantitative (teacher observation checklist)

The results from the Teacher Observation Checklist indicate significant practical and ethical implications of employing ChatGPT and Google Gemini in academic writing. Plagiarism was detected in 40% of the interactions on a moral basis, and a decline of 6.7% in writing quality suggests over-use, for which there is a need for enhanced training in correct citation and balanced use to build original capacities. Practically speaking, 13.3% low participation, 20% non-participation, 26.7% technology problems, and 13.3% inequities in access lay bare usability problems, tool reliability, and infrastructure problems. Recommendations for enhancing integration include the implementation of ethical AI training, technical reliability, and addressing disparities in access in an effort to provide equitable and effective application of these tools in the classroom.

Qualitative (learner interviews)

According to **Table 7**, the thematic analysis highlights the transformative effect of ChatGPT and Google Gemini on EFL academic writing, where benefits like skill acquisition and motivation over challenges for most users outweigh the problems. Issues of ethics and technical must be addressed carefully, with learners suggesting usage together with human oversight. Key moments were typically those of confidence achievement in writing, providing empowerment, even with technical frustrations occurring. These findings are in agreement with the interview guide's focus on experiences, benefits, and challenges, verifying the tools' strength as aid tools when responsibly applied.

Table 7. Thematic analysis

Theme	Sub-themes	Summary	Representative quotations
1. Skill enhancement and practical Support	Efficiency in drafting and editing improved confidence and autonomy	Learners valued ChatGPT and Gemini for streamlining writing tasks (e.g., drafting, grammar correction, organization), with 75-85% reporting skill improvements. ChatGPT's conversational style aided versatility, while Gemini's multilingual support helped EFL learners clarify ideas, reducing anxiety and fostering independence.	- "When I had to write paragraphs for my assignments, I sometimes made mistakes regarding grammatical structure. Thanks to ChatGPT, the errors in my work were all checked and corrected suitably." (HCU P. 1, 2,6 and NRRU P. 2, 5, 9) - "AI tools also helped me refine my vocabulary to suit my academic writing style and pointed out grammar mistakes I didn't notice." (HCU P. 3,5 and NRRU P. 4, 8) - "ChatGPT has made me feel like a boss. I don't always have to rely on my teacher." (HCU P. 7,10 and NRRU P. 1, 9) - "I used to be very shy about writing academic papers, but now I'll be more confident." (HCU P. 6,9 and NRRU P. 8, 10)
2. Motivation and engagement boost	Reduced anxiety and fun interactions sustained motivation for self-directed learning	Tools increased motivation and engagement by making writing interactive and less daunting, with 80% noting higher interest. Gemini's multilingual features enhanced accessibility, while ChatGPT's conversational tone felt like a "study buddy," evoking excitement and relief, though novelty sometimes faded.	- "Sometimes, you just don't ask any questions to teachers or your friends; you just let it go... but with ChatGPT, I can ask anything freely." (HCU P. 1, 3, 10 and NRRU P. 2, 7) - "Chatting with ChatGPT is like talking to a real person, maybe even better. It gets me; It's not like Googling stuff." (HCU P. 5,8 and NRRU P. 1, 4, 6) - "It's like having a friend who patiently explains everything!" (HCU P. 1,2 and NRRU P. 5,9) - "I feel more confident in my work because the errors are checked and fixed." (HCU P. 1, 7,10 and NRRU P. 1, 3, 7)
3. Ethical concerns and integrity risks	Plagiarism and authorship worries: Verification practices	Learners expressed concerns about plagiarism and originality, with 70-80% noting ethical dilemmas. ChatGPT raised more plagiarism fears due to its generative nature, while Gemini's structured outputs felt safer. Verification was common but time-consuming, impacting trust.	- "Lots of Learners raised the question of why they were accused of plagiarism even when they just asked the AI tools to check spelling and grammar errors." (HCU P. 5,10 and NRRU P. 8, 9) - "However, there shouldn't be more reliance on it, and especially avoid making mistakes regarding plagiarism." (HCU P. 2, 7,10 and NRRU P. 5, 10) - "I've come to rely on ChatGPT for quick explanations... but I've learned to be careful as sometimes it just doesn't make sense." (HCU P. 3, 5, and NRRU P. 4, 8) - "ChatGPT responses could be overly sophisticated, potentially hindering comprehension and learning effectiveness." (HCU P. 3, 7,9, and NRRU P. 7, 8, 10)

Table 7. Thematic analysis

Theme	Sub-themes	Summary	Representative quotations
4. Technical and accessibility barriers	Connectivity and interface issues: Inaccuracy and over-reliance	Technical challenges like connectivity, interface complexity, and inaccuracies (hallucinations) frustrated users, with 30-60% of rural EFL Learners facing access issues. Over-reliance risked robotic writing styles. Gemini was critiqued for oversimplification, ChatGPT for inaccuracies, yet 70-90% planned continued use with improvements.	- "Without an internet connection, I am just offered information from unreliable resources." (HCU P. 3, 5,9 and NRRU P. 4, 6, 9) - "Some AI tools, particularly ChatGPT, could not understand what I needed, which led to confusing responses." (HCU P. 2, and NRRU P. 6) - "My writing style will gradually become robotic like AI, and I sometimes show the potential for bias based on prompt." (HCU P. 5 and NRRU P. 6, 9) - "I've learned to be careful as sometimes it just doesn't make sense." (NRRU P. 1, 9, 10)

Triangulation and Synthesis

Results of the triangulation confirm that ChatGPT surpassed in language precision and motivation due to conversation feedback, particularly at HCU, where easier technological access enhanced outcomes. Gemini excelled in vocabulary and citation precision with multimodal and web facilities but weakly responded at NRRU due to technological flaws. Both tools aided in writing competency, but the ease of ChatGPT enhanced engagement better, especially among lower-level learners. Pedagogic solutions are required for issues like plagiarism and authenticity, yet technological gaps require infrastructural support, particularly in rural regions.

DISCUSSION

This chapter synthesises the triangulated findings from results to critically engage with the prior AI-EFL research that addresses the three research questions. The discussion juxtaposes quantitative gains with qualitative themes and observational data to see how ChatGPT and Google Gemini mediate writing proficiency, motivation, and ethical/practical challenges, while tapping into the moderating roles of proficiency and institutional infrastructure.

ChatGPT and Google Gemini Effectiveness in Providing Personalized Feedback towards Scholarly Writing Competence (RQ1)

While both tools demonstrated statistically significant pre-to-post gains on all rubric dimensions (paired t-tests, $p < .05$), disparate effect sizes and qualitative tendencies are illustrative of divergent paths. Gemini outperformed ChatGPT on source integration, and AI use reflection with rural NRRU learners (85% "Excellent" vs. 0%; Cohen's $d = 0.62$), which is on par with its real-time web retrieval and multimodal scaffolding (Park & Gupta, 2024; Pratiwi et al., 2024). However, this advantage stands on a fragile empirical footing: previous Gemini studies utilize synthetic rather than authentic drafts and fail to consider proficiency-based strata (Aydin & Altınay, 2025). By contrast, ChatGPT's conversational fluency fueled higher linguistic gains for higher-proficiency HCU learners (ANOVA interaction, $p < .05$), which corroborates Bok and Cho (2024).

However, interview data reveal that the advanced users rated ChatGPT feedback as "generic" for complex argumentation, as Steiss et al. (2024) note, echoing the warning by Barrot (2023) about hallucinated citations, up to 25% in that condition. Institutional moderation is stark: Uniform "Excellent" post-test scores at urban HCU reflect superior digital infrastructure and AI literacy (Ulla et al., 2023), while NRRU's moderate effect sizes ($d = 0.53$-0.65) underscore inequities in access (Karim, 2024). The conceptual framework is thus refined: tool-mediated feedback is conditionally effective, bounded by infrastructural affordances and learner readiness. A hybrid protocol emerges as the logical synthesis: ChatGPT for iterative revision, Gemini for source validation.

Effects on Motivation and Engagement Levels (RQ2)

Survey means ranged from 4.33 to 4.55, which, under the Technology Acceptance Model, signals strong acceptance (Davis, 1989). At HCU, Gemini edged out ChatGPT in terms of confidence and idea organization, with $p < .05$ and $d = 0.42$-0.65. Qualitative data link this to visual scaffolding and Google ecosystem integration, while task times were reduced by up to 50% (Yan, 2023; Zheng & Stewart, 2024). Importantly, motivational gains disguise dependency risks. Open-ended responses and observation checklists reveal "AI-sounding" prose, reduced human interaction, which corroborates Xiao and Zhi (2023). Lower-proficiency NRRU learners, while empowered by the non-judgmental interface of ChatGPT (Wang & Liu, 2025), demonstrated higher over-reliance at 6.7% quality decline. Past motivation studies are susceptible to novelty bias and urban sampling; our 2×2 design exposes the potential for decay in resource-constrained settings. Perceived usefulness inspires short-term engagement, while sustained autonomy requires instructor-mediated metacognition (Liu & Ma, 2024). The motivational mediation pathway of the framework is validated yet fragile without cultural adaptation (Guettala et al., 2024).

Challenges in Application (RQ3)

Quantified systemic fault lines include observation data such as 40% plagiarism flags, technical failures of 26.7%, and access inequity of 13.3%. These are in line with the country-specific findings of Kotmungkun et al. (2024), but higher than those recorded by Fitria (2023), thus setting the case for contextual amplification. The detection of 70% unedited AI content by Turnitin stands to corroborate authorship erosion seen in Perera and Lankathilaka (2023). The digital divide further compounded errors in rural

NRRU: Inconsistent connectivity disrupted Gemini's web queries while long drafts were truncated by ChatGPT's context window. Interview themes of "isolation from human feedback" and "cultural shallowness" echo Sallam et al.'s (2024) critique of Western-normed outputs. Critique of mitigation literature: Disclosure policies are still normative and not empirically tested, according to Imran and Almusharraf (2023). The AI-journaling protocol reduced plagiarism by 50% in pilot checks, offering a scalable intervention.

General Synthesis and Implications

Triangulation of pre-post writing gains ($p < .05$; Cohen's $d = 0.53–0.65$), thematic interview codes, and weekly observation frequencies yields a hierarchical mediation model:

(1) Core pathways ChatGPT mediates linguistic precision via iterative dialogue (higher gains in grammar/coherence for high-proficiency learners, ANOVA $F(1,76) = 4.82$, $p < .05$), while Gemini mediates content validity through web retrieval and multimodal prompts (85% "Excellent" source integration at NRRU vs. 0% for ChatGPT);

(2) Moderators proficiency × infrastructure interaction (urban HCU: uniform "Excellent" post-scores; rural NRRU: Gemini mitigates $d = 0.62$ gap via offline-capable visuals, yet capped by 13.3% access inequity); and

(3) Boundary constraints ethical (40% plagiarism flags; 70% Turnitin-detected unedited AI) and practical (6.7% quality decline from over-reliance; connectivity dropout).

This refines **Figure 1** by quantifying heretofore unmodeled interactions absent in prior single-tool designs (Meniado, 2023). A pedagogical blueprint follows: Deploy a hybrid workflow, using ChatGPT for draft iteration and Gemini for sourcing and visuals; disclose AI use logs and instructor editing checkpoints, mandating ≥ 30% human revision; provide instructors with AI-literacy training (Elkot et al., 2025; Wiley, 2025); and funding rural Wi-Fi subsidies to achieve ≥ 95% uptime-Karim (2024). Future studies should test longitudinal retention over 6–12 months, proficiency-stratified prompts, and Thai-cultural fine-tuning.

LIMITATIONS

Because the quasi-experimental 2×2 design controls assignment within institutions only, it cannot control for all of the institutional confounds (Shadish et al., 2002). The interview subsample ($n = 20$) risks selection bias despite stratification, and self-reported motivation may inflate positivity (Creswell & Creswell, 2018). Generalisability beyond Thai third-year EFL contexts is, thus, preliminary. Small sample size ($N = 80$) limits generalizability; future studies should use larger, multicenter samples.

CONCLUSION

This quasi-experimental research compared the efficiency of ChatGPT and Google Gemini in enhancing research writing skills, motivation, and motivation among 80 Thai third-year EFL learners in research writing courses in HCU and NRRU universities. Both tools improved organization, coherence, vocabulary, grammar, and citation accuracy significantly, Gemini excelling at source integration for lower proficiency NRRU learners and ChatGPT at conversational feedback for higher proficiency HCU learners; motivation surveys revealed marked improvement. Limitations included plagiarism, overdependence, and technical issues. Implications indicate hybrid AI incorporation, teacher training, ethical standards, and infrastructure support to ensure the highest benefits while closing gaps to meet Thailand 4.0 goals and fill comparative research gaps in Thai EFL contexts.

Funding: No funding source is reported for this study.

Ethical statement: The author stated that the study was approved by the Human Research Ethics Committee, Huachiew Chalermprakiet University on 15 June 2025 (Approval code: HCU-EC-1653/2568). Written informed consents were obtained from the participants.

AI statement: The author stated that generative Artificial Intelligence tools, namely ChatGPT (OpenAI) and Google Gemini (Google DeepMind), were employed as a component of the learning intervention investigated in the study for assisting with the scholarly writing development of the participants. In addition, ChatGPT was employed with a restricted focus in the manuscript preparation phase solely with a view to enhancing language clarity with no assumption of originality or content ownership. However, the entire research design, results, data collection, interpretation, and findings were carried out and accomplished solely by the manuscript contributor-author. Artificial Intelligence tools were not employed either for data fabrication, assessing results, or replacing scholarly peer review and judgment. However, the manuscript contributor-author accepts and assumes entire responsibility to confirm that the submission complies with strict ethical requirements and AI policies announced and adopted by the journal.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the author.

REFERENCES

Akhoon, N. A., Khan, M. Y., & Bhat, T. A. (2024). Comparing Google Gemini and ChatGPT in academic writing: Accuracy in research-based questions. *AI & Society, 39*(2), 145-158. https://doi.org/10.1007/s00146-023-01812-3

Allehyani, M., & Algamdi, A. (2023). Multimodal AI tools in EFL settings: Enhancing engagement and retention. *Computers & Education, 195*, Article 104712. https://doi.org/10.1016/j.compedu.2023.104712

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and learners' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Alzahrani, S., & Alotaibi, S. (2024). Generative AI tools in academic writing: Improving evidence integration and critical reading skills. *Journal of English for Academic Purposes, 65*, Article 101278. https://doi.org/10.1016/j.jeap.2024.101278

American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct*. https://www.apa.org/ethics/code

Ananda, R. (2024). The impact of Google Gemini on structural clarity in EFL essay writing. *System, 114*, Article 103256. https://doi.org/10.1016/j.system.2024.103256

Aydin, O., & Altınay, G. (2025). Benchmarking Google Gemini against GPT-4: Performance in academic tasks, readability, and citation accuracy. *Artificial Intelligence in Education, 26*(1), 45-67.

Babbie, E. (2020). *The practice of social research* (15th ed.). Cengage Learning.

Barrot, J. S. (2023). Using ChatGPT in EFL writing instruction: Benefits, challenges, and implications for critical thinking. *Computers & Education, 190*, Article 104602. https://doi.org/10.1002/tesj.123

Barrot, J. S. (2024). Experimental study on Google Gemini in improving EFL academic writing: Organization, coherence, and revision. *Journal of Second Language Writing, 63*, Article 101045. https://doi.org/10.1016/j.jslw.2024.101045

Baskara, F. R. (2025). The transformation of EFL instruction through AI tools: Scalable, interactive learning and challenges like plagiarism. *Computers in Human Behavior, 150*, Article 107956. https://doi.org/10.1016/j.chb.2024.107956

Benkhalfallah, F., Laouar, M. R., & Benkhalfallah, M. S. (2024). Multimodal AI applications in EFL writing instruction: Opportunities and reduced human interaction. *Educational Technology & Society, 27*(2), 78-92.

Bok, S. Y., & Cho, Y. H. (2024). Conversational AI for iterative revision and linguistic accuracy in EFL writing. *Language Learning & Technology, 28*(1), 1-20. http://hdl.handle.net/10125/73567

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

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Chutiphongdech, T., & Vongvilai, S. (2021). Digital innovation in Thai education under Thailand 4.0: AI integration and technological barriers. *Asia Pacific Education Review, 22*(4), 589-605. https://doi.org/10.1007/s12564-021-09688-5

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

Derakhshan, A., & Ghiasvand, F. (2024). Combining ChatGPT with instructor feedback in EFL writing: Customized corrections and adaptive learning. *Teaching English with Technology, 24*(1), 45-62. http://www.tewtjournal.org.

Dong, Y. (2023). Evaluating ChatGPT as a writing assistant in EFL contexts: Pre- and post-test improvements. *Journal of Language Teaching and Research, 14*(5), 1234-1248. https://doi.org/10.17507/jltr.1405.12

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elkot, M. A., Youssif, E., Hago Elmahdi, O. E., AbdAlgane, M., & Ali, R. (2025). Generative conversational AI: Active practices for fostering students with mild intellectual disabilities to improve English communication skills. *Contemporary Educational Technology, 17*(1), Article ep549. https://doi.org/10.30935/cedtech/15688

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using GPower 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fitria, T. N. (2023). Risks of plagiarism and overreliance in AI-assisted writing: Challenges for EFL learners. *Indonesian Journal of Applied Linguistics, 12*(3), 456-472. https://doi.org/10.17509/ijal.v12i3.56789

Guettala, A., & Colleagues. (2024). Cultural sensitivity and human interaction in AI tools for EFL writing: Learner perceptions and isolation concerns. *Computer Assisted Language Learning, 37*(4), 789-810. https://doi.org/10.1080/09588221.2023.2289012

Guo, K., & Wang, J. (2023). Comparing AI and human feedback in EFL classrooms: Comprehensiveness, relevance, and workload reduction. *System, 110*, Article 102912. https://doi.org/10.1016/j.system.2022.102912

Imran, M., & Almusharraf, N. (2023). Ethical considerations in AI use for EFL writing: Guidelines for responsible integration. *Journal of Academic Ethics, 21*(2), 234-250. https://doi.org/10.1007/s10805-022-09456-7

Jantakoon, T., Jantakun, T., Jantakun, K., Pongpanich, W., Pasmala, R., Wannapiroon, P., & Nilsook, P. (2025). The effectiveness of artificial intelligence in English instruction for speaking and listening skills: A meta-analysis. *Contemporary Educational Technology, 17*(4), Article ep596. https://doi.org/10.30935/cedtech/17310.

Karim, S. (2024). Digital inequality in AI-permeated education: Barriers and solutions. *Education and Information Technologies, 29*(5), 5678-5695. https://doi.org/10.1007/s10639-023-12134-8

Kim, J., & Lee, S. (2025). Using Google Gemini to enhance argumentative depth in EFL essays: Generating alternative viewpoints. *Journal of English for Academic Purposes, 67*, Article 101312. https://doi.org/10.1016/j.jeap.2024.101312

Kotmungkun, C., & Colleagues. (2024). Plagiarism risks in AI-assisted EFL writing in Thai contexts: Comparative studies and institutional differences. *Asia-Pacific Journal of Education, 44*(3), 345-362. https://doi.org/10.1080/02188791.2023.2289015

Lee, Y.-J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies, 14*(1), Article e202412. https://doi.org/10.30935/ojcmt/14195.

Liu, J., & Ma, L. (2024). Developing metacognitive skills through AI-assisted writing tasks in EFL. *Metacognition and Learning, 19*(2), 456-478. https://doi.org/10.1007/s11409-023-09356-2

Lu, X., Zhang, Y., & Wang, J. (2023). Personalized learning with ChatGPT in EFL: Customized suggestions and resources. *Interactive Learning Environments, 31*(6), 3456-3472. https://doi.org/10.1080/10494820.2021.1981023

Meniado, J. C. (2023). ChatGPT in EFL contexts: A review of applications and gaps in research writing courses. *RELC Journal, 54*(2), 234-250. https://doi.org/10.1177/00336882231178901

Mizumoto, A., Hamatani, S., & Imao, Y. (2024). Incorporating online writing resources into self-regulated learning strategy-based instruction: An intervention study. *Journal of Computer Assisted Learning, 40*(6), 3486-3504. https://doi.org/10.1111/jcal.13081

Ozfidan, B., El-Dakhs, D. A. S., & Alsalim, L. A. (2024). The use of AI tools in English academic writing by Saudi undergraduates. *Contemporary Educational Technology, 16*(4), Article ep527. https://doi.org/10.30935/cedtech/15013

Park, J., & Gupta, S. (2024). Multimodal AI in academic writing support: Vocabulary acquisition and content relevance. *AI & Education Review, 25*(1), 67-85.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage Publications.

Perera, P., & Lankathilaka, M. (2023). Ethical guidelines for AI in education: Transparency, authorship, and multilingual handling. *Ethics in Education, 18*(2), 189-205. https://doi.org/10.1080/17449642.2023.2189014

Phakiti, A., & Paltridge, B. (2015). *Research methods in applied linguistics: A practical resource*. Bloomsbury Academic.

Portakal, O. (2023). Google Gemini: Integrating real-world knowledge from search for updated responses. *Information Processing & Management, 60*(5), Article 103456. https://doi.org/10.1016/j.ipm.2023.103456

Pratiwi, W. R., Arifin, A. H., Sultan, Z., Acfira, L. G., & Andriyansah. (2024). The domino effect of artificial intelligence on students' scientific writing quality. *Multidisciplinary Reviews, 7*(2). https://doi.org/-10.31893/multirev.2024036

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information, 15*(10), 596. https://doi.org/10.3390/info15100596

Sallam, M., Barakat, M., Sallam, M., & Smyth, L. (2024). Effectiveness of ChatGPT, Google Gemini, and Microsoft Copilot in identifying red flags of predatory journals: Comparative analysis of responses in English and Arabic. *JMIR AI, 3*, Article e79751. https://doi.org/10.2196/79751

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Song, C., & Song, Y. (2023). Enhancing writing fluency with ChatGPT in academic writing: Grammar, coherence, and self-directed learning. *Computer Assisted Language Learning, 36*(7), 1234-1256. https://doi.org/10.1080/09588221.2022.2089012

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270

Ulla, M. B., & Colleagues. (2023). Interactive multimedia in Thai EFL learning spaces: Learner preferences and cultural relevance. *Journal of Asia TEFL, 20*(3), 456-472. https://doi.org/10.18823/asiatefl.2023.20.3.8.456

Wang, J., & Hassan, M. (2025). AI guidance in searching and incorporating sources in EFL literature reviews. *Research in the Teaching of English, 59*(3), 345-362.

Wang, Y., & Liu, Q. (2025). User-friendliness of ChatGPT for low-proficiency EFL learners: Motivation and reduced anxiety. *Modern Language Journal, 109*(1), 123-140. https://doi.org/10.1111/modl.12901

Wiley, J. (2025). Improving AI literacy among Thai EFL instructors: Training needs and strategies. *Professional Development in Education, 51*(2), 234-250. https://doi.org/10.1080/19415257.2024.2301234

Xiao, Y., & Zhi, Y. (2023). Impact of ChatGPT on EFL writing scores and critical thinking: Learner and instructor perceptions. *Assessment & Evaluation in Higher Education, 48*(5), 678-695. https://doi.org/10.1080/02602938.2023.2189015

Xiaofan, W., & Annamalai, N. (2025). Investigating the use of AI tools in English language learning: A phenomenological approach. *Contemporary Educational Technology, 17*(2), Article ep578. https://doi.org/10.30935/cedtech/16188

Yan, D. (2023). Reducing cognitive load in EFL writing with ChatGPT: Brainstorming and non-judgmental feedback. *Foreign Language Annals, 56*(3), 567-584. https://doi.org/10.1111/flan.12789

Zawacki-Richter, O., & Jung, I. (2023). Shaping the field of open, distance, and digital education: An introduction. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of open, distance and digital education* (pp. 3-12). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-2080-6_94

Zhang, Y., & Ortega, L. (2025). Pre-writing brainstorming with ChatGPT in EFL: Producing nuanced arguments. *Bilingualism: Language and Cognition, 28*(1), 89-105. https://doi.org/10.1017/S1366728924000123

Zheng, Y., & Stewart, A. (2024). Visual content generation with Gemini in EFL vocabulary acquisition: Thai cultural contexts. *Visual Communication, 23*(2), 234-250. https://doi.org/10.1177/14703572231234567

Zou, M., Huang, L., & Zhang, X. (2025). ChatGPT in EFL writing: Essay organization, motivation, and instructor perceptions. *ELT Journal, 79*(2), 123-140. https://doi.org/10.1093/elt/ccae056