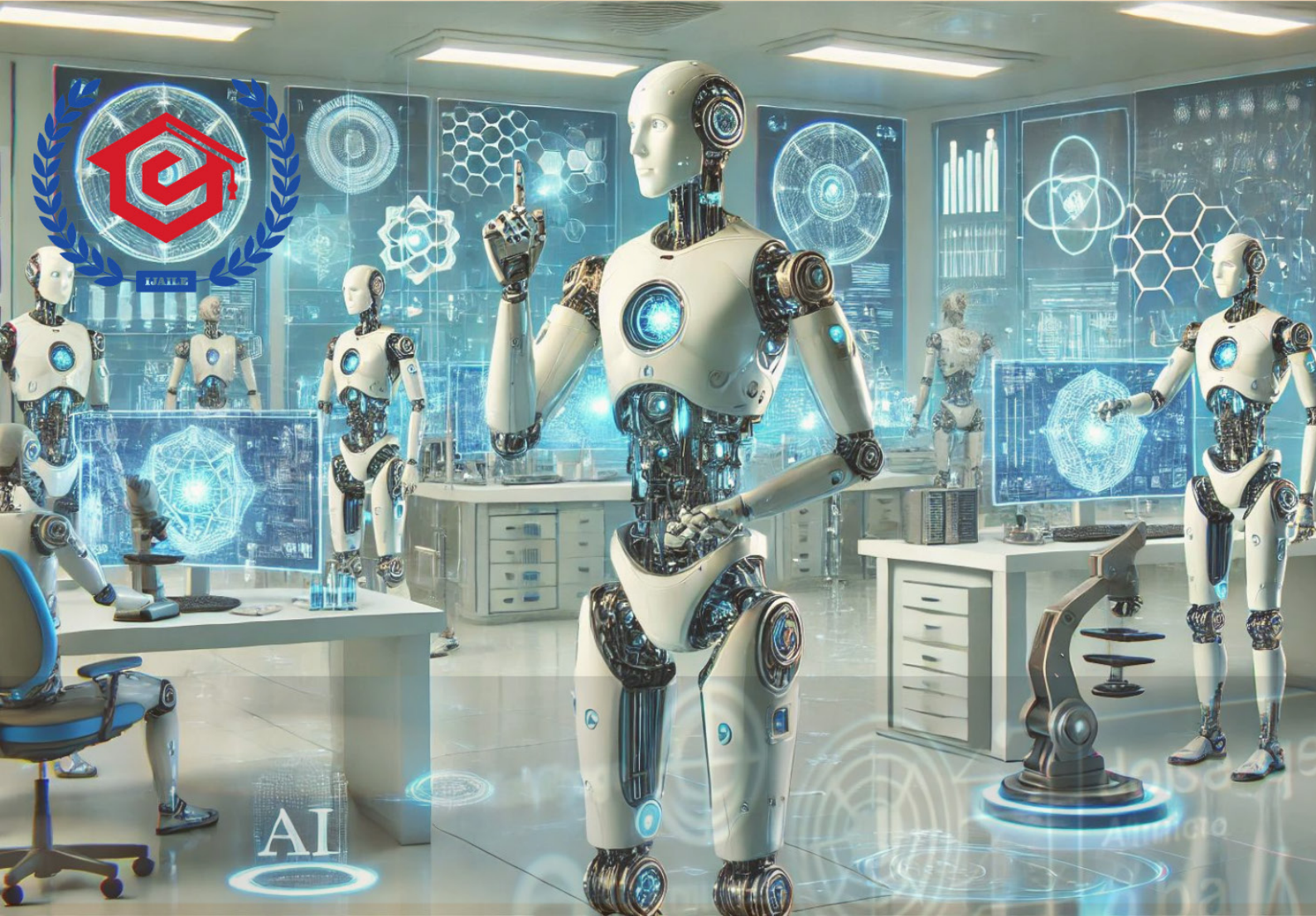




International Journal of **AI in Language Education**

ISSN: 3065-0216



ICTE Press

International Association of TESOL & Education

📍 5243 Birch Falls LN, Sugar Land, Texas, USA, 77479

🌐 <https://ijaile.org> | <https://i-cte.org/press>

International Journal of AI in Language Education

ISSN: 3065-0216

Volume 3, Issue 2, 2026



Indexed

Crossref
Google Scholar
LOCKSS
CLOCKSS
PKP PN
Library of Congress
U.S. ISSN Center
ROAD
Semantic Scholar
ResearchGate



Publication frequency

The IJAILE (ISSN: 3065-0216) will publish 3 issues per year in April, August, and December. The IJAILE will publish manuscripts as soon as they are accepted. The journal volume is defined by the year of publication of articles.

Copyright(c) 2026 International Journal of AI in Language Education is published under a Creative Commons Attribution 4.0 International License.

Authors retain copyright and grant the journal the right of first publication with the work simultaneously licensed under a Creative Commons Attribution 4.0 International License that allows others to share the work with an acknowledgment of the work's authorship and initial publication in this journal.

Publisher: ICTE Press

Address: International Association of TESOL & Education, 5243 Birch Falls Ln, Sugar Land, Texas, USA, 77479

Principal Contact

Associate Professor Dr. Pham Vu Phi Ho
Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam
Email: editor@ijaile.org or editor.ijaile@gmail.com

International Journal of AI in Language Education

Vol. 3, No. 2 (2026): AI in language Education

Editorial Board

Editor-in-Chief

Professor Dr. Jeremy White, Ritsumeikan University, Japan ([Scopus ID](#)).

Associate Professor Dr. Vu Phi Ho Pham, Industrial University of Ho Chi Minh City, Vietnam ([Orcid](#)) ([Scopus](#)), ([ResearchID](#))

Editorial Team

Professor Dr. Hayo Reinders, Anaheim University, USA. He is also Professor of Applied Linguistics (Adjunct) at King Mongkut's University of Technology, Thonburi, Thailand ([Orcid](#)) ([Scopus](#))

Professor Dr. Glenn Stockwell, Associate Dean, Waseda University, Japan ([Orcid](#)) ([Scopus](#))

Professor Dr. M. Rafael Salaberry, Rice University, USA ([Orcid](#)) ([Scopus](#))

Assoc. Prof. Dr. Nguyen Huu Cuong, Van Lang University, Ho Chi Minh City, Vietnam ([Orcid](#)) ([Scopus](#))

Assoc. Prof. Dr. Pham Thi Huong, Ho Chi Minh City University of Education, Vietnam ([Orcid](#)) ([Scopus](#)) ([ResearcherID](#))

Professor Dr. Andrew Lian, Suranaree University of Technology, Thailand; University of Canberra, Australia ([Orcid](#)) ([Scopus](#)).

Dr. Ania Lian, Charles Darwin University, Australia ([Orcid](#)) ([Scopus](#))

Associate Professor Dr. Sandro Barros, Department of Teacher Education, Michigan State University, USA ([Orcid](#)) ([Scopus](#))

Assoc. Prof. Dr. Nguyen V. Long, University of Foreign Language Studies, The University of Da-nang, Vietnam ([Orcid](#)) ([Scopus](#))

Assoc. Prof. Dr. Nguyen Ngoc Vu, Ho Chi Minh City University of Foreign Languages and Information Technology, Vietnam ([Orcid](#)) ([Scopus](#))

Dr. Thi Thuy Loan Nguyen, VASS College of Vocational Education, Australia, ([Orcid](#)) ([Scopus](#)) ([ResearchID](#))

Dr. Bao Dat, Monash University, Australia ([Orcid](#)) ([Scopus](#))

Dr. Hyangeun Ji, Department of Learning Technologies in the College of Information, University of North Texas, Texas, USA ([Scopus](#)), ([Google Scholar](#))

Dr. Mohd Norazmi bin Nordin, Faculty of Education, Universiti Kebangsaan, Malaysia ([Orcid](#)) ([Scopus](#)), ([Google Scholar](#))

Table of Content

Vol. 3, No. 2 (2026): AI in Language Education

Orcid: <https://orcid.org/0009-0004-5901-9176>

DOI: <https://doi.org/10.54855/ijaile.2632>

Research Articles

Pages

1. Generative AI Mediation in EFL Oral and Written Production: Colombian Student and Teacher Perception 1-18
Dionelio Jesús Moreno Villalobos (Author)
DOI: <https://doi.org/10.54855/ijaile.26321>
2. Leveraging Generative AI To Foster Teachers' Creativity in Lesson Design 19-32
Vu Thanh Loan, Nguyen Minh Tri (Author)
DOI: <https://doi.org/10.54855/ijaile.26322>
3. The Effects of ICTE Robot on Students' English-Speaking Skills: A Mixed-Methods Study on Non-English Majors at Nguyen Tat Thanh University 33-54
Pham Thi Ngoan, Pham Vu Phi Ho (Author)
DOI: <https://doi.org/10.54855/ijaile.26323>
4. Effectiveness of ChatGPT Feedback on Low-Level English Students' Writing Proficiency and Their Perceptions: An Experimental Study 55-70
Vu Quynh Hoa, Nguyen Kim Phuc (Author)
DOI: <https://doi.org/10.54855/ijaile.26324>



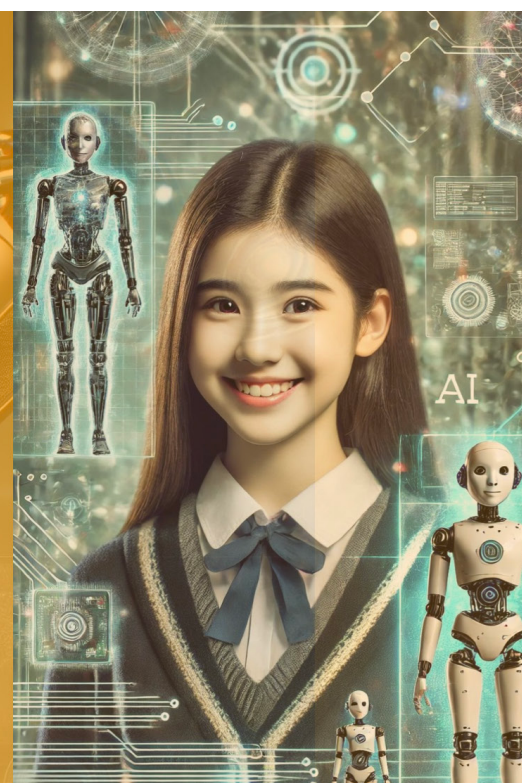
International Journal of
AI in Language Education

ISSN: 3065-0216



ICTE Press

International Association of TESOL & Education
5243 Birch Falls Ln, Sugar Land, Texas, USA, 77479
<https://ijaile.org> | <https://i-cte.org/press>



A Note from the Editor-in-Chief

Dear beloved Language Educators,

The International Journal of AI in Language Education (IJAILE) is pleased to announce the completion and publication of Volume 3, Number 2 (2026). This issue brings together four empirical studies examining the rapidly developing role of generative artificial intelligence and AI-powered tools in language education. Collectively, the articles address AI-supported speaking and writing, teacher creativity and lesson design, learner agency, formative feedback, and responsible human–AI interaction. The studies presented in this issue are situated primarily in EFL contexts in Colombia and Vietnam, offering important empirical insights into how students and teachers are beginning to integrate generative AI into everyday language learning and teaching practices.

In this issue, Villalobos (2026) explored how students and teachers perceived GenAI-supported oral and written production in a Colombian university language program. Seven students completed open-ended semi-structured questionnaires, while five teachers participated in semi-structured interviews. Using a qualitative exploratory design with a phenomenological orientation and thematic analysis, the study identified four interconnected patterns: GenAI as a rehearsal and revision partner, tensions between polished output and language ownership, the changing role of teacher mediation, and the need for clearer ethical and assessment guidance. The findings suggest that GenAI can support vocabulary development, rehearsal, revision, and confidence, but its pedagogical value depends substantially on learner agency, transparency, human feedback, and responsible authorship. A particularly valuable contribution of this study is its treatment of GenAI not simply as a writing tool but as a mediational resource connecting both oral and written production. The findings also emphasize that more polished AI-supported language does not necessarily indicate stronger independent language competence.

Vu and Nguyen (2026) investigated how generative AI influenced EFL teachers' creativity during a six-week professional development program at Nguyen Van Cu High School in Hai Phong, Vietnam. Ten in-service English teachers participated. The researchers adopted a mixed-methods design involving pre- and post-intervention lesson plans, a creativity assessment rubric, reflective journals, and semi-structured interviews. The results demonstrated improvements in originality, variety of instructional methods, multimodal integration, and learner orientation, with particularly strong gains in multimodal lesson design. Teachers increasingly perceived AI as an ideation partner and creative collaborator rather than merely an efficiency tool. At the same time, they expressed concerns about authenticity, cultural appropriateness, contextual relevance, and excessive dependence on AI-generated materials. The study therefore positions GenAI as a potential co-designer in lesson planning while maintaining that professional judgment, contextual adaptation, and pedagogical agency must remain with teachers.

Pham and Pham (2026) examined the impact of the customized ICTE Robot on the English-speaking performance of 24 non-English-major students at Nguyen Tat Thanh University. The mixed-methods study combined pre- and post-speaking tests adapted from the PET speaking rubric with semi-structured interviews involving nine students. Speaking performance was evaluated in vocabulary, grammar, pronunciation, and discourse management. The quantitative findings showed statistically significant improvement in vocabulary, pronunciation, and discourse management, whereas grammar did not improve significantly. Qualitative findings indicated that students valued the chatbot's voice-based and multimodal functions and reported enhanced confidence, vocabulary retention, pronunciation practice, flexibility, and immediate feedback. However, speech-recognition errors, microphone problems, connectivity issues, and inaccurate feedback were also reported. The study is particularly relevant because it examines a customized AI-supported learning environment rather than relying solely on widely available commercial chatbots. It also illustrates how voice-based AI may extend opportunities for speaking practice beyond conventional classroom interaction.

A Note from the Editor-in-Chief

Vu and Nguyen (2026) investigated the effectiveness of ChatGPT-generated formative feedback for low-proficiency, A2-level non-English-major university students. Forty-two students participated, with 21 students assigned to an experimental group receiving ChatGPT feedback and 21 to a control group receiving peer feedback and general teacher guidance. Pre- and post-writing tests were supplemented by semi-structured interviews. Students receiving ChatGPT feedback demonstrated significantly stronger writing outcomes than the comparison group. Qualitative findings further indicated benefits for grammar, vocabulary, content organization, immediacy, and personalized feedback. However, students also experienced inaccurate or overly general feedback, difficulties in crafting effective prompts, and risks of accepting AI-generated responses without sufficient evaluation. An especially noteworthy contribution concerns prompting behavior. Higher-performing students tended to use more structured prompting procedures, whereas lower-performing learners were more likely to rely on trial-and-error prompting. This finding raises important questions concerning the relationship among language proficiency, AI literacy, prompting competence, and learners' ability to benefit from automated feedback.

Overall, the four studies provide encouraging evidence that generative AI can support language education through speaking practice, writing feedback, teacher creativity, lesson planning, learner confidence, multimodal learning, and autonomous practice. However, their findings also show that effective AI use depends on how learners and teachers evaluate, adapt, and contextualize AI-generated output.

Several research gaps emerge. First, future studies should move beyond pre- and post-test outcomes to examine human–AI interaction processes through screen recordings, interaction logs, prompt histories, stimulated recall, and think-aloud protocols. Second, longitudinal research is needed to determine whether AI-supported gains lead to durable language development rather than temporary performance improvement. Third, controlled or crossover designs could compare AI-only, teacher-only, peer, and combined human–AI feedback to better isolate AI effects.

Further research should also investigate AI literacy and prompting competence, particularly how learners of different proficiency levels interpret and evaluate AI feedback. Teacher–AI co-creativity should be examined as an evolving process through classroom observations, design-based research, and lesson-planning analyses. Finally, research should explore human–AI complementarity, emphasizing models in which AI feedback, teacher expertise, learner reflection, and transparent assessment practices work together to support meaningful, responsible, and sustainable language learning.

The Editorial Team sincerely acknowledges and appreciates the scholarly contributions of all authors represented in this issue. Our gratitude goes to Dionelio Jesús Moreno Villalobos, Universidad EAN, Bogotá, Colombia, for contributing an important Colombian perspective on GenAI mediation, learner agency, authorship, and teacher guidance in EFL oral and written production.

We sincerely thank Vu Thanh Loan, FPT University, Vietnam, and Nguyen Minh Tri, Hoa Sen University, Vietnam, for their contribution to understanding how generative AI may support teacher creativity and the development of innovative, multimodal, and learner-centered lesson designs.

We also express our appreciation to Ngoan Thi Pham, Nguyen Tat Thanh University, Vietnam, and Vu Phi Ho Pham, Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam, for their empirical investigation of the ICTE Robot and its potential contribution to AI-supported English-speaking development.

Our sincere thanks are also extended to Vu Quynh Hoa and Nguyen Kim Phuc, Faculty of Foreign Languages, Hanoi University of Science and Technology, Vietnam, for their experimental research on ChatGPT-generated formative feedback for low-proficiency English learners and their examination of prompting practices and human–AI feedback integration.

A Note from the Editor-in-Chief

Their work collectively strengthens the evidence base for responsible, pedagogically grounded, and context-sensitive applications of artificial intelligence in language education.

Finally, the completion of Volume 3, Number 2 (2026) would not have been possible without the commitment of our reviewers, editors, editorial staff, and supporting colleagues. The International Journal of AI in Language Education sincerely thanks all peer reviewers who generously contributed their expertise and time to evaluating the submitted manuscripts. Their careful comments, methodological observations, constructive criticism, and recommendations have played an essential role in strengthening the academic quality, clarity, and rigor of the articles published in this issue.

We are equally grateful to the editors, editorial board members, and editorial staff for their sustained work throughout manuscript screening, peer-review coordination, revision, editing, reference checking, proof-reading, formatting, and production. Their dedication helps maintain the scholarly and ethical standards of the journal.

Call for Papers – Upcoming Issue

With the completion of Vol. 3, No. 2 (2026), the International Journal of AI in Language Education warmly invites researchers, language educators, teacher educators, graduate students, and interdisciplinary scholars to submit manuscripts for the upcoming issue. We welcome original empirical studies, mixed-methods research, experimental and quasi-experimental studies, qualitative investigations, systematic reviews, meta-analyses, design-based research, and theoretically grounded studies addressing emerging developments in artificial intelligence and language education.

Particularly encouraged are studies examining generative AI and language learning, AI-supported speaking and writing, AI-mediated assessment and feedback, AI literacy and prompting strategies, human–AI collaboration, teacher–AI co-creativity, learner agency and authorship, AI-supported teacher professional development, multimodal and conversational AI, ethical and responsible AI use, academic integrity, and innovative research methodologies for investigating human–AI interaction.

As artificial intelligence continues to reshape language education, IJAILE remains committed to publishing rigorous research that moves beyond technological novelty to examine how, why, when, and under what pedagogical conditions AI contributes meaningfully to language learning and teaching.

We look forward to receiving new submissions and to continuing our collaboration with researchers and educators worldwide in advancing responsible, evidence-based, and human-centered AI in language education.

Thanks God for everything!

Warm regards,
Editor-in-chief



Prof. Dr. Jeremy White



Assoc. Prof. Dr. Phan Vu Phi Ho
Editor-in-Chief

International Journal of AI in Language Education (IJAILE)



Generative AI Mediation in EFL Oral and Written Production: Colombian Student and Teacher Perceptions

Dionelio Jesús Moreno Villalobos^{1*}

¹Universidad EAN, Bogotá, Colombia

*Corresponding author's email: Djmorenov@universidadean.edu.co

 <https://orcid.org/0009-0006-8046-1123>

 <https://doi.org/10.54855/ijaile.26321>

® Copyright (c) 2026 Dionelio Jesús Moreno Villalobos

Received: 01/05/2025

Revision: 01/08/2026

Accepted: 11/08/2026

Online: 13/08/2026

ABSTRACT

Keywords: generative artificial intelligence; EFL production; teacher mediation; learner agency; AI literacy

Generative artificial intelligence (GenAI) is increasingly used to support EFL writing and speaking, yet research has focused more heavily on written products than on the combined processes of oral and written production, authorship, and teacher mediation. This qualitative study explored how students and teachers in a Colombian university language program perceived GenAI-supported production. Participants were seven students who completed an open-ended semi-structured questionnaire and five teachers who participated in semi-structured interviews. Data were examined through iterative thematic coding, analytic memoing, and comparison across participant groups. Four patterns emerged: GenAI as a rehearsal and revision partner; tension between polished output and language ownership; teacher mediation shifting toward critical language awareness; and the need for explicit ethical and assessment guidance. The study highlights that GenAI is pedagogically valuable when transparent, process-oriented, and dialogic routines preserve learner agency, human feedback, and responsible authorship.

Introduction

Generative artificial intelligence has changed the ordinary ecology of language learning. Students no longer approach writing and speaking tasks with only textbooks, dictionaries, classmates, and teacher feedback. They can now ask a chatbot to outline a paragraph, reformulate a sentence, explain an error, produce examples at a given proficiency level, simulate an interview, or prepare possible answers before an oral assessment. Recent studies have shown that GenAI tools can provide language examples, immediate feedback, conversational practice, and writing support, but they also create risks related to overreliance,

uneven feedback quality, inappropriate output, privacy, authorship, and assessment validity (Barrot, 2023; Kohnke et al., 2023; Warschauer et al., 2023; Yang & Li, 2024).

This development is especially relevant in Colombian university EFL contexts. Regional evidence indicates that AI tools are entering higher education and EFL practice rapidly, while their educational effects remain shaped by local conditions, institutional cultures, and unevenly developed guidance (George-Reyes et al., 2025; Gortaire-Díaz et al., 2025; UNESCO, 2023). Within the program examined here, English is linked to academic, professional, and international mobility, and students may use GenAI for planning, translation, drafting, rehearsal, feedback, and revision before teachers and programs have established shared criteria for acceptable assistance, transparency, authorship, and assessment. This mismatch creates the pedagogical gap examined in the study.

For EFL teachers, therefore, the central question is not simply whether students use GenAI. The more pedagogically important question is what happens to learning when a tool participates in the production of language that will later be assessed as evidence of student competence. Oral and written production are not merely products; they are mediated processes through which learners develop communicative competence, strategic control, voice, audience awareness, interactional responsiveness, and responsibility for meaning (Canale & Swain, 1980; Hymes, 1972; Lantolf & Thorne, 2006). When GenAI intervenes in those processes, it may support learning, but it may also obscure who made particular linguistic, rhetorical, or pragmatic decisions (Godwin-Jones, 2024; Warschauer et al., 2023). A polished paragraph or fluent oral script may therefore represent genuine learning, partial assistance, or substitution of learner agency.

The research gap addressed here is threefold. First, reviews and empirical studies show that AI-assisted writing and automated feedback dominate the literature, while speaking and the relationship between oral and written preparation remain less developed (Barrot, 2023; Crompton et al., 2024; Yan, 2024). Second, studies often emphasize affordances, attitudes, or product improvement, leaving authorship, language ownership, and teacher mediation less directly examined as interrelated processes (Kohnke et al., 2023; Pham & Le, 2024; Warschauer et al., 2023). Third, cross-national evidence has begun to include Latin American participants, but locally situated EFL studies from the region remain comparatively limited (George-Reyes et al., 2025; Gortaire-Díaz et al., 2025). The present study addresses these gaps by examining student and teacher perceptions of GenAI across connected oral and written production practices in a Colombian university program.

This article reports a qualitative study situated in a Colombian university language program. It does not ask whether GenAI should be accepted or prohibited. Instead, it asks how students and teachers perceive GenAI as a mediational resource in EFL oral and written production and what classroom routines may support transparent, responsible, and pedagogically meaningful use. The argument developed here is that GenAI becomes educationally meaningful only when its use is made visible, discussed, and connected to process-oriented assessment and teacher mediation. Without such mediation, the tool may improve the appearance of student output while weakening reflection, authorship, spontaneous performance, and human feedback.

Literature Review

GenAI as a Mediational Resource in Language Learning

This study is grounded in a sociocultural understanding of learning, where tools, signs, interaction, and social activity mediate development (Lantolf & Thorne, 2006; Vygotsky, 1978). From this perspective, GenAI should not be understood as a neutral accessory added to pre-existing language tasks. Tools change the activity in which they participate. A learner who asks a chatbot to compare two ways of expressing an idea is not simply receiving a correction; the learner is entering a mediated interaction in which noticing, choice, and self-regulation may be reorganized.

This mediational framing helps avoid two extremes. The first is technological enthusiasm, which treats GenAI as an automatic path to better learning because it generates immediate language. The second is technological suspicion, which treats every use of GenAI as a shortcut or threat. A mediational perspective asks a more analytical question: What kind of learning activity does the tool make possible, and under what pedagogical conditions does that activity strengthen or weaken learner agency? The same AI-generated suggestion can serve as a resource for reflection if students compare, evaluate, and explain it, or it can become a substitute for learning if they accept it without understanding.

Recent work on distributed agency in AI-supported language education similarly suggests that learning is shaped by the interaction among learners, teachers, tools, tasks, policies, and assessment practices (Godwin-Jones, 2024). This is particularly important in EFL contexts, where learners may overvalue fluent output because it resembles expert language, even when they cannot fully explain the choices that produced it. GenAI, therefore, functions as a mediational resource with uncertain pedagogical effects: it can scaffold production, but it can also narrow learning if used primarily to avoid difficulty.

Writing, Automated Feedback, and the Risk of Surface Improvement

Writing has been one of the most visible areas of GenAI use in language education. Students can use chatbots to generate outlines, correct grammar, propose titles, simplify language, expand vocabulary, improve coherence, and rewrite paragraphs. Research on ChatGPT and L2 writing has reported both possibilities and concerns. Barrot (2023) discusses the potential of ChatGPT for second language writing while warning about overreliance, uneven feedback quality, and the possibility that students may submit polished texts without developing stronger independent control. Escalante et al. (2023), Guo and Wang (2024), and Yan (2024) similarly show that AI-generated feedback can be useful, but only when learners actively process it, and teachers help them interpret the suggestions. More recent empirical work sharpens this distinction. In a double-blind, paired-comparison study, Tsai et al. (2024) found that ChatGPT-assisted revisions raised EFL essay scores, with the largest gains for initially weaker writers, yet cautioned that such revisions do not reveal learners' own writing competence and therefore call for new forms of writing assessment. In a multiple-case study, Yan and Zhang (2024) likewise showed that L2 writers' engagement with ChatGPT feedback varied with proficiency and strategy use, and that learners valued the tool partly because it reduced the discomfort of receiving critical feedback from humans. Read together, these studies indicate that AI-supported writing can improve products and lower affective barriers while leaving open whether independent competence actually develops—the precise tension this study examines in the EFL production process.

This concern is also reflected in recent IJAILE research on AI-supported academic writing. Tran, Le, and Tran (2025) reported that AI tools can support academic writing development

by improving skills and providing assistance, while also creating challenges related to over-reliance and limited support for complex writing problems. Their findings are relevant to the present study because they show that the value of AI support lies not only in the improved final product but also in whether learners understand and appropriate the revisions made.

The distinction between surface improvement and learning is crucial. When a text becomes more fluent after AI support, teachers may find it harder to evaluate the student's actual writing development. The final product may be cleaner, but it may not show stronger planning, organization, editing ability, or ownership of language choices. For this reason, the present study treats AI-supported writing not as a detection problem but as a mediation problem: how students use the tool, what they learn from it, and what they can still explain without it become central pedagogical questions.

Oral Production, Rehearsal, and Communicative Agency

A systematic review of artificial intelligence in English language teaching and learning has mapped affordances across speaking, writing, reading, pedagogy, and self-regulation, alongside recurring challenges such as technological breakdowns, limited tool capabilities, and the standardization of learner language (Crompton et al., 2024). Within this body of work, however, writing remains considerably better documented than speaking. Although writing has received more attention, GenAI also matters for oral production. Learners can use chatbots to prepare answers, create role-play scripts, translate possible expressions, generate vocabulary lists, simulate interviews, and rehearse before speaking in class. As voice-based interfaces become more common, learners may increasingly use AI not only for written preparation but also for spoken interaction. These practices may reduce anxiety and provide additional practice, especially for students who have limited opportunities to speak English outside the classroom.

Recent empirical studies on AI-mediated speaking support this point while also revealing its limits. Nguyen and Le (2025) found that the AI conversational agent, Call Annie, increased students' confidence and willingness to communicate in English-language lessons, although concerns about chatbot-mediated interaction remained. In a Latin American EFL setting, Gortaire-Díaz et al. (2025) likewise found that AI tools were perceived as accessible and helpful for reducing error-related anxiety in classroom language practice. These findings suggest that AI-mediated rehearsal can have affective value, but they do not resolve whether confidence developed with AI transfers to spontaneous human interaction.

Communicative competence provides a useful lens for this concern. Canale and Swain (1980) and Hymes (1972) remind us that language production involves more than formal accuracy. Learners need to make choices appropriate to context, purpose, audience, and interaction. In oral production, they must also manage hesitation, repair, turn-taking, clarification, and unpredictability. GenAI can help students prepare language, but it cannot guarantee that learners can use that language flexibly. Teachers therefore need to design oral tasks that allow AI-supported preparation while still requiring personal response, interaction, and justification.

Voice, Authorship, and Language Ownership

The concept of voice is central to this study because EFL production is not simply a matter of producing correct language. Learners gradually develop ways of positioning themselves, expressing attitudes, making choices, and sounding like legitimate participants in a second or foreign language. GenAI complicates this process because it can generate language that is more fluent, more academic, or more idiomatic than the learner's current repertoire. The result may be impressive, but it may also feel detached from the learner's own intention.

Authorship in AI-supported language production should not be understood as a binary category. A learner may remain meaningfully involved when using AI for planning, prompting, selecting, rejecting, revising, and explaining. However, authorship becomes weaker when the learner cannot account for the decisions that shaped the final version. In this sense, authorship is not only a legal or academic integrity category; it is also a pedagogical category related to language ownership and learner agency.

Pham and Le (2024) found that students perceived ChatGPT as useful for language learning, engagement, and knowledge development, while also warning that over-reliance may limit critical thinking and creativity. This ambivalent pattern is important for the present study because it suggests that GenAI can support learner voice only when students are guided to adapt, question, and appropriate the tool's output rather than treat it as authoritative language.

Digital Literacy, Ethics, and Pedagogical Transparency

Digital literacy in the GenAI era involves more than technical ability. This understanding extends the multiliteracies tradition, which reframes literacy as a plural, socially situated practice shaped by changing technologies and communicative contexts (New London Group, 1996). Students may know how to use a chatbot, but they may not know how to evaluate the quality of its response, identify hallucinated information, recognize inappropriate register, avoid overdependence, or document assistance ethically. In language education, these concerns are intensified because the object of learning is language itself. If a tool generates the language that students are expected to develop, then digital literacy and language pedagogy become inseparable.

Research on AI literacy emphasizes the need for users to understand, evaluate, and critically engage with AI systems (Ng et al., 2021). UNESCO (2023) likewise calls for human-centered, transparent, and responsible use of generative AI in education. For EFL programs, these principles imply that AI policies should not remain abstract institutional statements. They need to be translated into classroom-level agreements and assessment practices that clarify which forms of assistance are permitted, which require disclosure, and which undermine the learning objective.

A Latin American perspective on higher education further complicates this issue. George-Reyes et al. (2025) show that students across Latin America and Europe perceive ChatGPT as useful but complex in its educational impact, reinforcing the need to examine local institutional cultures rather than assume that global technologies are adopted in identical ways. In Colombian EFL classrooms, responsible GenAI use therefore requires attention to access, teacher workload, assessment expectations, and students' developing academic and professional language trajectories.

Research Questions

This study was guided by the following main research question:

How do students and teachers perceive the role of GenAI in EFL oral and written production, particularly in relation to revision, rehearsal, authorship, teacher mediation, assessment, and responsible use?

The analysis also considered three supporting questions:

- How do students describe their use of GenAI when preparing written and oral production tasks?

- How do teachers interpret the pedagogical, ethical, and assessment-related implications of GenAI use?
- What classroom routines may support transparent and responsible GenAI use in EFL production?

Methods

Research Design and Context

The inquiry follows a qualitative exploratory design with a phenomenological orientation. This orientation was chosen because the study seeks to understand how participants experience and give meaning to GenAI in relation to language learning, rather than to measure the effect of a specific intervention. The study does not claim that GenAI improves or harms EFL production in general. Instead, it examines how students and teachers interpret the tool in relation to learning, assessment, authorship, and pedagogy.

The study is situated in a private Colombian university language program where learners complete communicative tasks, written assignments, oral presentations, and classroom interactions in English. The Colombian EFL context is relevant because students often link English learning to academic performance, professional mobility, and intercultural participation, while teachers must respond to institutional assessment demands and rapidly evolving student practices. GenAI enters this environment not as an isolated innovation but as part of a broader transformation in how students access language support.

Participants and Sampling

A purposeful, non-probabilistic sampling strategy guided participant selection, consistent with the study's qualitative and phenomenological orientation. The focal corpus included 12 information-rich participants: seven students from an English-language and culture course in a modern languages program, aged 17 to 23, and five teachers affiliated with the same program, aged 30 to 55. Participants were selected because they had direct experience with GenAI tools in oral and written production and were willing to participate.

The small sample size is acknowledged as a limitation in breadth, but it is methodologically appropriate for this exploratory qualitative study. The aim was not statistical representativeness, but analytic depth: to identify how participants make sense of GenAI use, where tensions emerge, and how student and teacher perspectives converge or diverge. In this sense, the findings are intended as analytically transferable insights rather than generalizable claims about all Colombian EFL learners or teachers.

Table 1.

Participant Profile for the Article Corpus

Participant group	n	Age range	Primary data source	Role in analysis
Students	7	17-23	Open-ended semi-structured questionnaire through Forms	Accounts of use, benefits, concerns, ownership, and transparency
Teachers	5	30-55	Semi-structured in-depth interviews	Interpretations of mediation, assessment, feedback, and ethical guidance
Total	12	17-55	Qualitative perception data	Focal corpus for thematic analysis

Data Collection

Data collection followed two clearly differentiated phases. First, the seven student participants completed an open-ended semi-structured questionnaire through Forms. The student questionnaire asked about uses, benefits, concerns, and experiences with AI-supported oral and written production. It included prompts about planning, vocabulary, grammar correction, speaking rehearsal, confidence, dependence, teacher guidance, disclosure, and perceived risks.

Second, the five teacher participants completed semi-structured interviews. These interviews did not recruit a smaller subgroup from the student questionnaire; rather, they constituted the teacher component of the corpus. The interviews explored GenAI in relation to feedback, authorship, dependence, learner confidence, teacher mediation, assessment, and ethical responsibility. The semi-structured protocol included guiding questions such as: How do you perceive students' use of GenAI in oral and written production? What forms of AI assistance appear pedagogically acceptable or problematic? How does GenAI affect feedback, authorship, and assessment? What kind of teacher guidance is needed for responsible use?

Students, therefore, provided open-ended questionnaire accounts, while teachers provided interview data. The two sources were compared to identify convergences and contrasts between learners' experiences and teachers' interpretations. Documentary materials provided contextual support, but participant perception data remained the primary empirical basis of the article. To make the evidence traceable while protecting anonymity, the teacher interview excerpt is identified with a T code, whereas student questionnaire excerpts are identified sequentially as SR1, SR2, and so forth. The SR labels distinguish the quoted response excerpts used in the analysis rather than asserting participant-level links beyond the anonymized evidence presented here.

Data Analysis and Trustworthiness

The analysis used iterative thematic coding supported by analytic memos. Initially, the researcher read the data several times to identify recurrent meanings related to revision, rehearsal, voice, authorship, autonomy, feedback, mediation, dependence, assessment, and responsible use. Preliminary codes were then grouped into broader interpretive categories. Following Braun and Clarke's (2006, 2021) approach to thematic analysis, the process

involved moving between coded segments, emerging themes, participant accounts, and the overall research question.

Trustworthiness was strengthened through four procedures. First, the analysis compared student and teacher perspectives as a form of source triangulation. Second, dated analytic memos documented decisions about coding and theme formation and recorded the boundary maintained between descriptive findings and interpretive or pedagogical claims. Third, disconfirming and ambivalent cases were retained in the findings rather than removed; for example, some students described GenAI as useful support, while others warned that it could produce language beyond their real level. Fourth, an audit trail was maintained through dated coding notes, successive versions of the thematic tables, and a documented record of analytic decisions.

Because the study is qualitative and interpretive, reliability was not treated as inter-rater statistical agreement. Instead, methodological rigor was pursued through transparency of procedures, reflexivity, thick description of context, attention to divergent cases, and clear linkage between participant excerpts and interpretive claims (Lincoln & Guba, 1985). The researcher's position as a language educator provided contextual sensitivity but also required reflexive caution so that the analysis did not simply confirm prior pedagogical expectations.

Results/Findings

The analysis points to four interrelated patterns that help explain how GenAI is experienced in EFL production tasks. The findings below are organized to foreground empirical evidence, with interpretive and pedagogical claims developed in the Discussion and Pedagogical Implications sections. Each pattern is supported by student and teacher accounts, including divergent views and tensions within the corpus.

Across student accounts, GenAI most frequently appeared as a resource for organizing ideas, expanding vocabulary, correcting language, preparing for oral tasks, and reducing uncertainty before performance. Teachers, however, interpreted these benefits alongside concerns about authorship, over-polished output, and the loss of human feedback. This contrast is important: students tended to emphasize usefulness and confidence, while teachers tended to emphasize mediation, assessment, and the need for transparent criteria.

Table 2. Evidence-Based Thematic Structure

Theme	Evidence from participants	Tension identified in the data	Analytical meaning
GenAI as rehearsal and revision partner	Students described using AI to organize ideas, find vocabulary, rehearse answers, and receive immediate feedback.	Support lowered anxiety but could also reduce productive struggle if accepted passively.	GenAI functioned as a low-stakes preparation space whose value depended on reflective use.
Polished output versus language ownership	Some students valued clearer language; others noted that AI could produce language beyond their real level.	A better-looking text or script did not always indicate stronger independent control.	Authorship became a process of selecting, adapting, rejecting, and explaining AI output.

Teacher mediation as critical language awareness	Teachers emphasized that AI cannot replace the affective and formative dimension of human feedback.	AI feedback was fast, but human mediation was needed to interpret appropriateness, level, and purpose.	Teacher work shifted from correction alone toward judgement, transparency, and critical language awareness.
Need for ethical and assessment guidance	Participants expressed uncertainty about disclosure, acceptable support, and how AI-supported performance should be assessed.	Unclear rules encouraged private judgement and possible concealment.	Responsible use required explicit classroom agreements and process evidence.

GenAI as a Low-Stakes Rehearsal and Revision Partner

The first pattern concerns the way students experience GenAI as a low-stakes partner for rehearsal and revision. Students described using GenAI at moments when production felt difficult to begin. In writing, they used it to organize ideas, improve vocabulary, compare possible phrasings, and revise grammar. In speaking, they used it to prepare answers, anticipate vocabulary, rehearse ideas, and reduce anxiety before participating in class or completing an oral task.

One student summarized this supportive role as follows:

"GenAI has helped me organize my ideas. I feel more confident when structuring a sentence, and I have learned vocabulary because it offers synonyms and examples for speaking in an interview. The feedback is immediate; sometimes it is not as personalized as teacher feedback, but it guides me a lot." (SR1)

A second response described a more selective use:

"I normally try to use it as a model without finishing - a draft or brainstorm from which I can take various ideas and build a new one." (SR2)

These accounts show that students did not necessarily experience AI as a final author; several framed it as a starting point for planning, vocabulary expansion, and initial organization.

The same pattern appeared in oral rehearsal. One student reported using AI as a base for a presentation and then adapting it:

"I take the answer as a base for how to begin the presentation; after that I keep building it, I tell the AI how I want it to be and then I do the final adaptation, in which I change words, structures, and so on, depending on what I want to achieve." (SR3)

This account suggests an iterative relationship between learner intention and AI-generated support rather than simple copy-and-paste use.

However, the pattern was not entirely positive. Students recognized that rehearsal support may lead to dependence when the generated version replaces the learner's own effort. This tension is important because GenAI appears to lower the threshold for participation, but the data do not show that this support automatically becomes deeper learning. It becomes pedagogically meaningful only when students adapt, explain, and retain control over the language they use.

Polished Output Does Not Always Mean Language Ownership

The second pattern concerns the tension between polished output and language ownership. Students appreciated clearer phrasing, more precise vocabulary, and improved organization, yet some also recognized that AI-supported production could exceed their real level of control. One student stated:

"In my case, there were times when the AI said things I would not know how to formulate on my own, demonstrating a level I did not really have and would not know how to demonstrate if I had to do it without AI." (SR4)

This account illustrates a level-displacement problem, in which AI-supported output can outpace the learner's independent competence.

Another student expressed a similar risk in relation to oral presentations:

"The main risk is that AI suggests vocabulary or grammatical structures that are too advanced or complex, that we have not yet seen in class. This makes the text lose its naturalness and the student fails to memorize or defend the oral presentation well." (SR5)

This quotation is important because it links advanced vocabulary, loss of naturalness, and inability to defend the performance. In other words, the problem is not only ethical but also communicative and pedagogical.

A divergent view also appeared in the data. One student argued that AI does not necessarily create a false level because the learner still has to perform:

"Structurally, maybe the AI is better than the student; but in expressing ideas, body language, and pronunciation, the student will always have a better level, and likewise in improvising on the spot." (SR6)

This disagreement suggests that AI-supported production should not be reduced to plagiarism or dependence. It may displace language ownership in some cases, but it may also support students when they remain able to adapt, pronounce, explain, and respond spontaneously.

These accounts show that authorship in AI-supported language production is negotiated through multiple decisions: what the student wants to say, how the student prompts the tool, which suggestions are accepted and which are rejected, how the final text or script is revised, and whether the student can explain the final choices. The evidence, therefore, supports a process-based understanding of ownership rather than a binary distinction between AI use and no AI use.

Teacher Mediation Shifts from Correction to Critical Language Awareness

The third pattern concerns the transformation of teacher mediation. Teacher accounts did not describe GenAI as making pedagogy irrelevant. Instead, they suggested that teachers must help learners interpret, evaluate, and contextualize AI support. One teacher summarized this concern as follows:

"There is more confidence for speaking in simulated contexts, but GenAI lacks the affective and formative dimension of human feedback." (T1)

This comment captures a central concern among teachers: AI may provide rapid language support, but it does not replace the relational, affective, and formative dimensions of teacher feedback.

Student accounts also confirmed the value of guidance. One student explained:

"It helped me that the teacher always made us see AI not as a problem but as a tool. Likewise, the teacher gave us ideas about how to ask the AI correctly - not so that it would do the work for us, but so that we could receive orientation in case of confusion." (SR7)

This excerpt shows that teacher mediation can reposition AI use from a hidden shortcut to a guided learning resource.

Another student described teacher-guided oral practice with a conversational interface:

"My English teacher recommended I use the ChatGPT avatar to practise as if I were having a real conversation or doing an exposition. This tool helped me improve my pronunciation, corrected mistakes, and suggested expressions that sounded more natural in English." (SR8)

This account suggests that teacher mediation can shape the purpose of AI use, especially when the tool is framed as rehearsal and feedback rather than as a substitute for performance.

The difference between student and teacher perspectives is therefore not a simple opposition. Students tended to value immediacy, confidence, and vocabulary support. Teachers tended to emphasize the need for human feedback, appropriateness, and responsibility. Together, these perspectives suggest that GenAI shifts teacher mediation from correction alone toward critical language awareness: helping learners decide whether AI-generated language is appropriate, understandable, level-consistent, and genuinely theirs.

Ethical and Assessment Guidance Is a Pedagogical Need

The fourth pattern concerns the need for explicit ethical and assessment guidance. Students did not speak about disclosure in a uniform way. Some viewed it as necessary; others treated it as dependent on the teacher or on how much of the AI output was used. One response linked disclosure to teacher reactions:

"It depends on the teacher, since some take it badly and would make an issue of it." (SR9)

A contrasting response framed unacknowledged use as acceptable when AI served only as inspiration:

"It is not necessary, as long as it is not exactly the same, but rather you take what it gave you as inspiration." (SR10)

These responses suggest that unclear rules lead students to make private decisions about what counts as legitimate use.

Other students articulated a stronger position on transparency. One stated:

"Yes, I consider it important to inform the teacher when artificial intelligence is used to prepare an oral activity. This promotes transparency and lets the teacher understand how the work was done. Moreover, using AI as support does not take merit away from learning when it is used responsibly." (SR11)

One respondent explained that disclosure helps "show what we learned or whether we have doubts about something the AI explained" (SR12).

These accounts link transparency not only to integrity but also to formative feedback.

Assessment tensions also arose when students considered whether AI could make a performance appear stronger than the learner's actual level:

One respondent warned that memorizing AI-generated language could lead to an "excellent" presentation without understanding what was being said (SR13).

Another observed that AI may introduce more technical vocabulary when asked to produce a script (SR14).

Together, these accounts support the claim that AI-mediated production creates a visibility problem: teachers may see the product but not the process by which it was produced.

The Colombian university setting matters here because students and teachers are negotiating the use of GenAI in a context where institutional expectations, course-level practices, and everyday student uses are still being aligned. The data do not support a simplistic ban, but neither do they support uncritical adoption. They instead point to the need for explicit agreements regarding acceptable support, disclosure, process evidence, and spontaneous demonstration of communicative control.

Discussion

The findings show that GenAI is best understood as a pedagogically unstable mediational resource. It can support EFL production by providing vocabulary, examples, rehearsal, feedback, and confidence, but it can also create dependence, level displacement, loss of ownership, and uncertainty about assessment. This interpretation extends previous research by showing that the central issue is not only whether AI improves output but also whether students remain able to explain, adapt, and perform in the language that AI helped them prepare.

The writing-related findings are similar to Barrot (2023), Guo and Wang (2024), Yan (2024), and Tran et al. (2025) in that they confirm the usefulness of AI feedback and revision support while also warning against overreliance. The difference is that the present study foregrounds language ownership as an interpretive and pedagogical issue. In this Colombian EFL corpus, students did not simply ask whether AI made writing better; they also questioned whether AI-generated language still represented their own level and voice. This adds a learner-ownership dimension to discussions often centered on feedback quality or writing improvement.

The oral-production findings also converge with Nguyen and Le (2025) and Gortaire-Díaz et al. (2025), who report that AI-supported practice may increase confidence and reduce anxiety. However, the present study qualifies that claim by showing that confidence may be rehearsal-bound. A student may feel safer after practicing with AI, but that confidence may not transfer to spontaneous interaction if the learner memorizes a script or uses language that cannot be defended. Thus, AI-mediated speaking should be evaluated not only through perceived confidence but also through spontaneous follow-up, explanation, and adaptation.

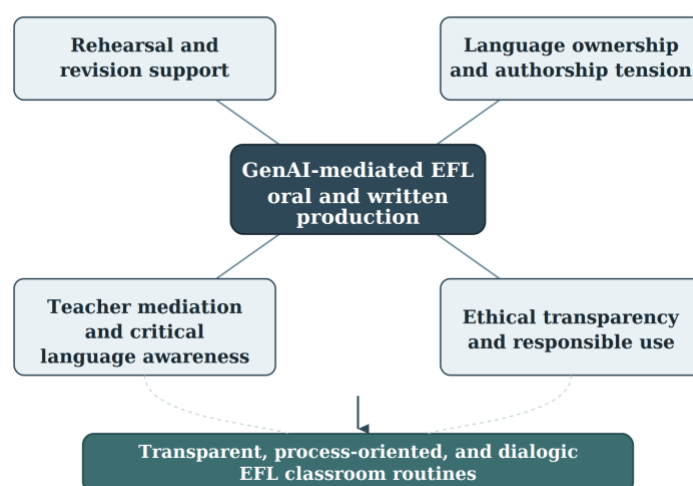
The findings further extend the work of Pham and Le (2024) and Kohnke et al. (2023) by showing that critical AI literacy is not an abstract competence. It appears in small classroom decisions: choosing vocabulary the student can actually use, rejecting language that sounds too advanced, disclosing AI assistance, or asking the teacher how to use AI without outsourcing the task. In this sense, teacher mediation does not restrict autonomy; it creates the conditions for critical autonomy. Without such mediation, students may be operationally independent but pedagogically dependent.

A counterargument must also be considered. GenAI may expand access to language support for students who lack opportunities for practice outside the classroom. It can provide immediate input, reduce embarrassment, and offer explanations at any time. From this perspective, overly restrictive AI use may reduce equitable access to support. Nevertheless, the negative evidence in this study suggests that access alone is insufficient. If AI-generated language is too advanced, if students cannot defend their performance, or if disclosure rules are unclear, the tool may produce a more fluent product while making learning less visible. The pedagogical task is therefore not to choose between access and restriction, but to design mediated access.

The study's unique contribution is its integrated view of oral and written production, teacher mediation, and authorship from a Colombian university EFL context. Rather than treating GenAI solely as a writing assistant, the findings show that students move between writing and speaking: they draft, translate, rehearse, script, revise, and prepare oral responses with the same tool. This connection means that AI-aware pedagogy must address productive skills together and design assessment practices that distinguish preparation from performance.

Figure 1.

Conceptual Model of Responsible GenAI Mediation in EFL Oral and Written Production



Note. The figure is read from the four surrounding dimensions inward to the central practice and downward to the outcome. The four dimensions correspond to the interpretive patterns reported in the findings.

The model synthesizes the four interpretive patterns identified in the findings. It places GenAI-mediated EFL oral and written production at the center, surrounded by the dimensions through which participants experienced the tool: rehearsal and revision support; the tension between polished output and language ownership; the repositioning of teacher mediation toward critical language awareness; and the need for ethical transparency and responsible use. The downward arrow indicates that these dimensions do not function in isolation. GenAI becomes pedagogically valuable only when its use is converted into transparent, process-oriented, and dialogic classroom routines through teacher guidance, learner reflection, and ethical transparency. In the absence of such mediation, the same tool may yield more polished output while leaving learning less visible.

Pedagogical Implications for EFL Programs

The implications below are derived from the findings rather than presented as general advice detached from the data. Because students reported both support and risk, EFL programs should avoid categorical responses that either ban GenAI completely or allow it without guidance.

First, language programs should teach the use of AI as a decision-making process. The student who described using AI as a draft or brainstorm illustrates a productive pattern: the learner begins with AI support but builds a personal version. Teachers can reinforce this by asking students to identify what they accepted, changed, rejected, and learned.

Second, the assessment should include evidence of process. The level-displacement accounts show that a polished final product may not represent independent competence. Process-oriented assessment can include initial drafts, AI-assisted revisions, oral defenses, spontaneous follow-up questions, short reflection notes, and teacher-student conferences.

Third, transparency should be normalized rather than treated only as a disciplinary matter. Several students linked disclosure to fear of teacher reaction, while others linked it to responsible learning. This means that students need clear, task-specific statements explaining when AI use is allowed, when it must be declared, and which parts of the task must be produced independently.

Fourth, teacher education should focus on AI-aware task design. Teachers need examples of prompts, rubrics, disclosure statements, revision activities, speaking tasks, and feedback routines. The findings suggest that teachers remain central because they help learners evaluate whether AI-generated language is appropriate for level, purpose, audience, and communicative context.

A Classroom Protocol for Responsible GenAI Use

The classroom protocol proposed here is grounded directly in the study's findings: students valued GenAI as support, warned against overdependence, and asked for clearer guidance about transparency and assessment. It is offered as a transferable framework rather than a prescriptive sequence, to be adapted to local proficiency levels, task types, and institutional norms.

Before submitting an AI-supported written or oral task, learners can complete four brief prompts: (a) I used GenAI to help me with...; (b) the most useful suggestion was...; (c) I accepted, changed, or rejected the suggestion because...; and (d) this is what I can now explain or perform without the tool. These prompts make the process visible without turning the assignment into surveillance.

For writing tasks, students can submit an initial draft, an AI-assisted revision, and a short commentary explaining changes. For speaking tasks, students can use GenAI during preparation but must answer at least one unscripted follow-up question, respond to a peer, or explain a vocabulary choice. For feedback tasks, students can compare teacher feedback and AI feedback, identify where they agree or differ, and justify which suggestion they used.

The protocol should be adapted to the proficiency level. Beginners may use simple disclosure frames such as "I used AI for vocabulary" or "I changed these words." More advanced students can write fuller reflective commentaries. The goal is not to punish support, but to protect the distinction between assistance, appropriation, and substitution.

Limitations and Future Directions

This study has limitations that should be stated explicitly. First, the sample is small ($n = 12$), which limits the breadth of perspectives represented. The findings should therefore be read as exploratory and analytically transferable rather than statistically generalizable. Second, the study is situated in a single Colombian university language program; institutional culture, teacher practices, student access to technology, and assessment norms may differ in other contexts.

Third, the study relies primarily on self-reported perceptions through student questionnaire responses and teacher interviews. These data are valuable for understanding meaning-making, but they do not directly measure learning outcomes or demonstrate actual improvement in oral or written production. Fourth, the study did not include classroom observation, AI interaction logs, prompt histories, written drafts, or recorded oral performances. As a result, the analysis cannot fully compare participants' reported behavior with their actual behavior during AI-supported production.

Future research should incorporate larger, more diverse participant groups, compare institutions, and include classroom observations, AI use logs, writing samples, prompt records, and oral performance data. Longitudinal designs would be especially valuable for examining whether AI-supported rehearsal and revision lead to durable language development or mainly improve immediate task performance. Future studies should also investigate Latin American EFL contexts more systematically, since current international evidence remains concentrated in other regions.

Conclusion

GenAI is already reshaping how learners approach EFL writing and speaking. The question for teachers is not whether the technology exists, but how language programs will respond pedagogically. The findings discussed in this article suggest that students and teachers are not simply divided between enthusiasm and resistance. They are negotiating a complex terrain in which GenAI can support revision, rehearsal, vocabulary growth, and confidence while also raising concerns about authorship, dependence, level displacement, assessment, and the loss of human feedback.

The unique contribution of this study is to show that GenAI-mediated EFL production should be understood as a matter of mediated ownership. Students may use AI responsibly when they select, adapt, explain, and disclose support. They may lose ownership when they accept language they cannot understand, defend, pronounce, or adapt. Teachers, therefore, remain essential as mediators of critical language awareness, process visibility, and ethical judgment.

For EFL programs, this means treating GenAI as a curricular issue rather than merely a disciplinary problem. Students need to learn how to use AI support without surrendering ownership of their language. Teachers need institutional support to design tasks that make the process visible. Programs need shared expectations that distinguish responsible assistance from substitution. If these conditions are present, GenAI may help learners develop greater awareness of language choices. If they are absent, the tool may simply produce better-looking texts and scripts while leaving learning less visible.

Compliance with Ethical Standards

Ethical considerations: The study followed institutional ethical principles for research involving human participants. Participants were informed about the purpose of the study, the

voluntary nature of their participation, and the confidential treatment of their responses. The use of open-ended questionnaires and interviews was aligned with the qualitative purpose of understanding perceptions rather than evaluating participant performance.

Informed consent: Informed consent was obtained from participants before data collection. Participants were informed that their contributions would be anonymized and used for academic research purposes.

Conflict of interest: The author declares no conflict of interest.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability: Due to the qualitative nature of the study and the need to protect participant confidentiality, the raw data are not publicly available. Anonymized excerpts may be reported in publications when they do not compromise participant identity.

Declaration of the Use of AI Technologies in the Writing Process

During the preparation of this manuscript, OpenAI was used only for language editing, clarity enhancement, organization of ideas, and editorial refinement. The author was responsible for the study design, theoretical framing, data interpretation, argument development, reference verification, and final approval of the manuscript. No AI tool was used to generate research data, conduct data analysis, fabricate references, or replace scholarly judgment. No AI tool is listed as an author.

References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47. <https://doi.org/10.1093/applin/1.1.1>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55(6), 2503-2529. <https://doi.org/10.1111/bjet.13460>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- George-Reyes, C. E., Avello-Martínez, R., & Buenestado-Fernández, M. (2025). Perceptions of ChatGPT and the complexity of its impact among higher education students: Evidence across ten countries of Latin America and Europe. *Educational Process: International Journal*, 15, Article e2025171. <https://doi.org/10.22521/edupij.2025.15.171>

- Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5-30. <https://doi.org/10.64152/10125/73570>
- Gortaire-Díaz, D., Almache-Granda, G., Real-Roby, R., & Mora-Herrera, E. (2025). Bridging the gap: How AI tools support self-directed EFL skill development inside the classroom. *Revista InveCom*, 6(2), Article e602012.
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Penguin.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Nguyen, L. A. D., & Le, T. T. P. (2025). Exploring the effects of an AI chatbot on emotional engagement in English speaking lessons: Insights from Call Annie. *International Journal of AI in Language Education*, 2(2), 79-99. <https://doi.org/10.54855/ijaile.25225>
- Pham, V. P. H., & Le, A. Q. (2024). ChatGPT in language learning: Perspectives from Vietnamese students in Vietnam and the USA. *International Journal of Language Instruction*, 3(2), 59-72. <https://doi.org/10.54855/ijli.24325>
- Tran, H. N., Le, T. T. N., & Tran, V. B. U. (2025). AI tools in learning academic writing: Benefits and challenges for MA students in the English language studies at the Industrial University of Ho Chi Minh City. *International Journal of AI in Language Education*, 2(1), 74-91. <https://doi.org/10.54855/ijaile.25215>
- Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 29(17), 22427-22445. <https://doi.org/10.1007/s10639-024-12722-y>
- UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or

- foreign language. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1-19. <https://doi.org/10.64152/10125/73597>
- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11, Article 1086. <https://doi.org/10.1057/s41599-024-03543-y>
- Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, 124, Article 103351. <https://doi.org/10.1016/j.system.2024.103351>

Biodata

Dionelio Jesús Moreno Villalobos is a university language professor and doctoral researcher in Education based in Bogotá, Colombia. His research focuses on generative artificial intelligence in language education, multilingualism, digital literacies, learner autonomy, and authentic assessment. He has experience in university language teaching, academic research, translation, and scholarly publishing.

Leveraging Generative AI To Foster Teachers' Creativity in Lesson Design

Vu Thanh Loan^{1*}, Nguyen Minh Tri²

¹ FPT University, Vietnam

² Hoasen University, Vietnam

*Corresponding author's email: loanvt4@fe.edu.vn

*  <https://orcid.org/0009-0003-7679-4046>

 <https://doi.org/10.54855/ijaile.26322>

® Copyright (c) 2026 Vu Thanh Loan, Nguyen Minh Tri

Received: 21/10/2025

Revision: 04/06/2026

Accepted: 13/07/2026

Online: 13/08/2026

ABSTRACT

Keywords: AI, CALL, EFL teachers, creativity, lesson design

Although generative artificial intelligence (GenAI) has gained popularity to assist in planning lessons in EFL teaching, its potential to enhance teachers' creativity remains underexplored. Thus, the study aims to investigate the impact of GenAI on EFL teachers' creativity in lesson design during a six-week professional development program at Nguyen Van Cu High School, Hai Phong, Vietnam. The study was carried out with ten in-service EFL teachers. A mixed-methods design was utilized, including pre- and post-intervention lesson plans, creativity evaluation rubrics, and reflective journals. Quantitative data were analyzed through creativity scores, whereas qualitative data were analyzed thematically. The results suggested that GenAI improved teachers' originality, multimodal lesson design, and learner-centered instructional practices. However, participants were worried about authenticity and overreliance on AI as well. The findings highlight GenAI as a valuable pedagogical partner for enhancing teacher creativity while underscoring the necessity for reflective and balanced integration of AI into language education.

Introduction

Generative artificial intelligence (AI) technologies have emerged, with tools such as ChatGPT, significantly influencing the landscape of education, particularly language teaching. These technologies enable educators to create educational content, customize learning materials, and maximize classroom interactivity. Previous studies have mainly examined student-centered applications of AI, which have been found to improve learning outcomes, engagement, and assessment (Kasneci et al., 2023; Zawacki-Richter et al., 2019). However, research examining teachers' use of generative AI in instructional design remains relatively limited compared with research focusing on students' learning experiences and outcomes.

Although previous research has focused on the productivity and feasibility of AI, there is still limited empirical evidence on how AI enables and constrains teachers in lesson planning, particularly in authentic classroom contexts. More specifically, prior studies have not

adequately investigated (1) how teachers integrate AI into lesson design practices (Hoang et al., 2025), (2) how creativity may be systematically measured during AI-assisted instructional design, and (3) how teachers' perceptions of creativity change as a result of continued use of AI tools. This discrepancy is particularly clear in EFL settings, where inflexible curricula and examination-oriented teaching practices constrain teachers' creative agency (Richards, 2013).

Creativity (Runco & Jaeger, 2012) is widely defined as the ability to generate original and appropriate ideas or artifacts in a given context and is emerging as a core competency for effective teaching. Creative educators develop standardized content into meaningful learning experiences, adapt instructional materials to learners' diverse needs, and design authentic, relevant learning activities. However, it remains difficult to sustain such creativity due to time constraints, institutional demands, and inadequate professional support.

Within this context, generative AI can be conceptualized as a cognitive and pedagogical support tool that enables idea generation, content reuse, and multimodal lesson planning. Rather than just automating routine tasks, AI has the potential to enhance teachers' creative processes by offering diverse perspectives, supporting instructional ideas, and facilitating rapid prototyping of teaching materials. The question is whether teachers use AI as a creative partner or primarily as a tool to improve efficiency. Addressing this distinction is important for understanding the educational value of AI integration and its potential contribution to teacher creativity.

Based on this gap in the research, the current study seeks to examine how EFL teachers use generative AI tools in lesson design and how such use influences their creativity. The study took place at Nguyen Van Cu High School in Hai Phong, Vietnam, where English teachers operate within a national curriculum framework that places considerable emphasis on standardized learning outcomes and examination preparation. Similar constraints have been documented in Vietnamese EFL contexts, where teachers often face heavy curricular requirements and limited flexibility for pedagogical innovation (Nguyen & Burns, 2017; Pillai et al., 2022). Through a six-week professional development program, this research examines how generative AI can function not merely as an instructional assistant but as a collaborative partner in lesson design. By integrating quantitative analysis of lesson-plan creativity with qualitative data from teachers' reflections, this research aims to provide nuanced insights into how AI is shaping teachers' creativity in contemporary EFL contexts.

Literature review

Generative AI in education

AI refers to machine learning systems that can automatically generate new outputs (text, images, teaching ideas, quizzes, multimedia) from human input and large-scale training data (Shanahan et al., 2023). Educational institutions have increasingly adopted this technology. Unlike previous generations of educational technology that did not rely on static databases or programmed responses, generative AI uses deep learning models (especially transformer architectures) to replicate creative, context-specific production (OpenAI, 2023). Such systems, including ChatGPT, Gemini, Claude, and DALL·E, are increasingly being applied in educational practice for writing support, adaptation, automated feedback and grading, and assessment creation. Recent research has pointed out a number of pedagogical affordances of generative AI. Kasneci et al. (2023) demonstrated that AI-based text generators can provide different levels of materials, which further scaffold differentiated instruction by building materials at scale, allowing teachers to match the needs of different learners more effectively.

AI has already been used in language learning to simplify reading passages, create vocabulary exercises, or translate genuine sources while preserving context. Similarly, teachers can use AI to form discussion prompts anchored in Bloom's taxonomy or conduct dialogues to assess communicative competence. However, when used effectively, it potentially helps foster teachers' ideational fluency and pedagogical imagination (Yang et al., 2025). For example, teachers can stimulate multimodal engagement among students by using AI to suggest interesting role-play scenarios. This perspective is in line with that of Holmes et al. (2019), who argue that these tools can help to reduce technical barriers for teachers and thereby provide greater opportunities for instructional innovation and inspiration. Despite this, the use of generative AI in education still faces issues of authenticity, authorship, and ethical use. Issues such as factual inaccuracies ("hallucinations"), risks of plagiarism, and the opacity of AI-generated explanations may undermine teachers' trust in the technology (Dwivedi et al., 2023). Additionally, empirical research remains limited on teachers' use of generative AI in the classroom, particularly regarding their creative decision-making. Recent investigations reveal that the effectiveness of AI in education relies primarily on teachers' critical literacy, flexible pedagogies, and institutional support (Luckin et al., 2024).

The role of creativity in language teaching

Teacher creativity means the capacity to develop innovative, useful, and contextually relevant methods of instruction that involve students and fulfill curriculum-driven purposes in teaching (Richards, 2013). In language teaching, creativity is exemplified by the creation of communicative tasks, the alteration of materials to accommodate each lesson, and the construction, from the start, of authentic learning experiences that invite students to explore through curiosity and imagination. Creativity operates on four dimensions (Kaufman & Beghetto, 2009): "mini-c" creativity denotes personally meaningful interpretations and insights; "little-c" creativity denotes everyday creative problem-solving; "Pro-C" creativity denotes professional-level expertise and innovation; and "Big-C" creativity includes eminent contributions that transform a field. This framework serves as a useful lens for understanding how teachers use creativity in lesson design at the "little-c" and "Pro-C" levels, where pedagogical adaptation and professional expertise are most evident in the context of this study. This study, therefore, operationalizes teacher creativity in terms of observable instructional outputs rather than abstract notions.

Many studies have shown that teacher creativity enhances student outcomes, and so does children's creativity. Research has consistently shown that creative teaching promotes intrinsic motivation, participation, and long-term learning (Richards, 2013; Sawyer, 2011). Likewise, Nguyen & Le (2025) demonstrated that students experience higher levels of engagement and communicative confidence when teachers use creative storytelling, gamification, or digital media. In the EFL context, creativity promotes adaptive teaching by enabling instructors to tailor lessons to students' linguistic and cultural backgrounds (Xie & Fang, 2025). The development of teachers' creativity comes from both individual and systematic factors. According to Craft (2005) and Sawyer (2011), teachers are discouraged from trying new ideas due to less flexible curricula, an exam-oriented culture, and pressures coming from the administration. As a result, creative activities are often neglected. On the other hand, schools that foster reflection, peer interaction, and continuous professional development (CPD) are more likely to promote teacher creativity (Jones & Richards, 2015). Creativity in language education, in particular, is not only about developing new materials, but about creatively recontextualizing previously familiar material in new forms. Creative teachers use authentic materials, develop role-plays on contemporary topics, or use digital materials to concretize abstract linguistic processes. As digital tools become increasingly integrated, the idea of "digital

creativity” has emerged, highlighting teachers’ capacity to utilize technology as a medium for pedagogical transformation rather than reducing it to a “mechanical one” to some extent (Mishra & Koehler, 2006). Such understanding provides a conceptual basis for understanding how generative AI might affect teachers’ creative processes in language classrooms.

Generative AI and teacher creativity

Generative AI and teacher creativity represent an emerging area of educational research. Some initial work suggests that AI might facilitate teachers’ ideation by generating multiple alternatives and visual inspiration, as well as help iterate on ideas (Holmes et al., 2019; Hossain & Al Younus, 2025). Teachers who use AI-powered programs, such as ChatGPT, Canva Magic Write, and Gemini, describe rising “idea fluency,” producing hundreds of new teaching ideas in a flash (Hossain & Al Younus, 2025). The tools can also serve as brainstorming tools, providing templates for lesson plans, storylines, and assessment tasks that teachers can modify and refashion within the pedagogical curriculum.

Furthermore, generative AI could stimulate divergent thought by introducing educators to cross-cultural or interdisciplinary perspectives (Yang et al., 2025). For example, a language teacher could leverage AI to produce idiomatic expressions from different English regions, or to write themed lessons in science, art, and social issues. This is an analogous dynamic within the creative paradigm of concept expansion, as new combinations of ideas arise when heterogeneous information sources are combined.

However, embedding AI in creative pedagogies also presents substantial challenges. Bender et al. (2021) and Kasneci et al. (2023) caution that uncritical reliance on AI-generated outputs may reduce users’ active engagement in creative processes and encourage more formulaic approaches to content creation and instructional design. For example, teachers may directly adopt AI-generated lesson plans or activities without adjusting them to the learner’s proficiency level, cultural context, or curriculum requirements. In EFL classrooms, this can lead to tasks that are linguistically inappropriate or pedagogically misaligned with learning objectives. Ethical concerns, including privacy and bias issues, as well as intellectual property issues, complicate the educational aspects of AI use (Williamson & Eynon, 2020).

For these reasons, in order to optimize benefits, researchers propose a “human-AI co-creativity” framework (Long & Magerko, 2020), so that AI is not a competitor to teacher skillfulness, but rather that it provides a vehicle for exploration, reflection, and development. Under this theory, teachers continue to be the principal facilitators of creativity, critically engage with AI outputs, and contextualize them for learners. Professional development programs oriented towards AI literacy, pedagogy, and ethics are thus important (Kasneci et al., 2023; Holmes et al., 2019).

With these tensions and opportunities, further empirical research is required to explore how teachers engage with generative AI in authentic educational settings. By studying the ways in which AI mediates the creativity of teachers - through its presence as a potential source of support, inspiration, or potential constraint - the research article contributes to theoretical and practical contributions to teacher education in the digital age. Therefore, the current study examines how high school EFL teachers at Nguyen Van Cu High School integrate generative AI tools into instructional design, and how it impacts their creative pedagogical practices.

Research questions

To fulfill the purpose of the study, the survey sought to answer the following research questions:

1. How do EFL teachers integrate generative AI tools into their lesson designs?
2. To what extent does the use of generative AI influence different dimensions of teachers' creativity in lesson design?
3. What changes occur in teachers' perceptions of creativity and AI-supported teaching after using generative AI tools?

Methods

Pedagogical setting & participants

The study adopted a mixed-methods design combining quantitative analysis of creativity scores with qualitative analysis of reflective journals. The intervention was structured over six weeks and involved the use of generative AI tools in EFL lesson planning. Ten in-service English teachers (8 female and 2 male) from Nguyen Van Cu High School in Hai Phong voluntarily enrolled. They had 3 to 15 years of teaching experience. All had basic digital literacy but limited experience with AI tools.

Purposive sampling was used to recruit participants. An invitation outlining the study's objectives, procedures, and expected commitments was issued to all English teachers at the school. The study was conducted with the participation of teachers who were willing and interested in integrating generative AI tools into their lesson-planning practices. Participation was completely voluntary, with no monetary or material benefits provided.

Despite the limited sample size ($N = 10$), this was reasonable given the exploratory nature of the study. No statistical generalization was expected; instead, we aimed to go beneath the surface of teachers' experiences, creative practices, and perceptions regarding the integration of generative AI into lesson planning. In exploratory educational research, small samples are often used to build rich contextual knowledge and new trends for large future research (Creswell & Plano Clark, 2018; Merriam & Tisdell, 2016).

Participant code	Gender	Years of experience	Prior AI experience
T1	Female	12	None
T2	Female	7	Minimal
T3	Male	10	None
T4	Female	3	Minimal
T5	Female	8	Moderate
T6	Female	15	None
T7	Male	9	None
T8	Female	6	Minimal
T9	Female	5	Minimal
T10	Female	11	Moderate

Design of the study

Participants completed a professional development program that targeted the generative AI application for lesson development.

- Weeks 1 to 2: Familiarization with AI tools (ChatGPT, Canva, and Gemini).

- Weeks 3 to 4: Workshops on lesson design with AI prompts – activity generators and creative adaptors.
- Weeks 5 to 6: Execution, review, and re-evaluation.

Each teacher provided a pre- and post-intervention lesson plan, which was evaluated using the Creativity Assessment Rubric (CAR) adapted from Richards (2013) and Kaufman & Beghetto (2009).

The instruments used in this study include:

- Creativity Assessment Rubric (CAR): Measured lesson plans across four dimensions: originality, variety of methods, multimodality, and learner orientation. Each criterion was rated from 1 (low) to 5 (high). To establish measurement validity, participants were required to provide both initial AI-generated outputs and their final lesson plans. Creativity scores were assigned based on the extent to which teachers modified, adapted, and contextualized AI-generated content in order to differentiate between teacher creativity and AI-generated input. To establish reliability, two independent raters reviewed all lesson plans. Inter-rater agreement was determined using Cohen's kappa, and the value of $\kappa = 0.82$ showed strong agreement. Discrepancies were resolved through discussion. The CAR, although originally designed to assess instructional products, is relevant to the present study because lesson plans directly represent teachers' pedagogical reasoning and creative decision-making, thereby providing a valid proxy for teacher creativity.
- Reflective Journals: Teachers wrote weekly reflections about the extent to which AI influenced their lesson design and imagination.
- Semi-structured Interviews: Semi-structured interviews were conducted after the intervention to further explore participants' perceptions and to clarify any issues or questions that arose from their teaching experience.

Data collection & analysis

To summarize the study, both qualitative and quantitative data were collected and analyzed. Quantitative data were analyzed using SPSS, using a paired-sample t-test to compare creativity scores of teachers before and after treatment. The results demonstrated a significant increase ($p < .001$). While the sample was small ($N = 10$), the use of parametric testing is well-accepted given the within-subject and exploratory nature of the study. For additional support in interpreting the analysis, effect sizes were estimated using Cohen's d , which provided information on the magnitude of the changes in size rather than statistical significance.

For the qualitative element, thematic analysis was conducted in accordance with the framework outlined by Braun and Clarke (2006). It was a step-by-step process. The data were coded using initial coding, which identified themes and then refined them. The final themes emerged from examining journals and interview data. They are AI as a source of inspiration; AI as a co-creator in lesson design; and concerns about AI-generated content being authentic. These themes together show how teachers used AI and how it helped them be creative with lesson plans. They interacted with AI in various ways. Their experience with AI shaped their lesson plans.

Findings

In this section, the results will be reported in response to three research questions, while also providing quantitative and qualitative analyses of the impact of generative AI tools on teachers' lesson design practices, creativity, and professional perceptions.

RQ1: How do EFL teachers integrate generative AI tools into their lesson designs?

Lesson plan review, reflective journaling, interviews, and use of data suggest that teachers interacted with generative AI at levels that are iterative, involving generation, adaptation, and transformation. At a fundamental level, AI was employed to generate ideas, such as brainstorming lesson themes, crafting prompts, and proposing classroom activities. Teachers used AI-generated outputs at a more advanced level, such as modifying reading texts, rewriting dialogues, or adjusting task instructions to better align with learners' skill levels and lesson objectives. At a higher level, some participants demonstrated a transformational use of AI outputs, which were critically evaluated, reorganized, and integrated into larger pedagogical designs. Importantly, AI tools were not deployed as standalone generators but were built into teachers' lesson-planning workflow alongside pre-lesson ideation, material development, and activity sequencing.

Participants gradually changed their methods (e.g., stopped using AI for linguistic support (e.g., checking grammar, translation) and began using AI as a creative collaborator in the design phase, especially in the interactive and learner-directed creation of activities). AI tools such as ChatGPT, Canva, and Copilot were increasingly used to inform multimodal lesson design by helping teachers visualize abstract concepts and augment instructional content.

RQ2: To what extent does the use of generative AI influence different dimensions of teachers' creativity in lesson design?

Data from pre- and post-tests were quantitatively analyzed using paired-sample t-tests in SPSS. Table 1 shows mean scores across four creativity dimensions before and after the intervention.

Table 1

Pre- and post-intervention creativity scores (N=10)

Dimension	Pre-test mean (SD)	Post-test mean (SD)	Mean difference	t(9)	p	d
Originality	2.8 (0.60)	4.3 (0.55)	+1.5	6.10	<.001	2.45
Variety of methods	3.1 (0.57)	4.4 (0.52)	+1.3	5.92	<.001	2.30
Multimodal integration	2.5 (0.67)	4.1 (0.54)	+1.6	6.44	<.001	2.60
Learner orientation	3.0 (0.70)	4.2 (0.63)	+1.2	5.03	=.001	1.80
Overall creativity	2.85 (0.64)	4.25 (0.54)	+1.4	6.17	.001	2.40

Note: $p < .001$ reflects statistical significance.

Most importantly, for the other four dimensions, all were statistically significantly improved. Among them, multimodal integration showed the largest expansion (Mean difference = +1.6), indicating that AI gave teachers the most support to effectively incorporate and use all kinds of semiotic materials in their lesson plans. Next in order was Originality (+1.5), which suggested a strengthened capacity to produce new instructional concepts. Learner orientation showed the smallest improvement (+1.2), suggesting that adapting instruction to learners' specific needs may still require substantial teacher judgment and pedagogical expertise. The general creativity score increased by +1.4. Effect sizes (Cohen's $d = 1.80-2.60$) were large across all dimensions, indicating substantial practical differences between pre- and post-intervention scores. Nevertheless, due to the limited sample size ($N = 10$), such findings are only preliminary and exploratory in nature, and not generalizable.

RQ3: What changes occur in teachers' perceptions of creativity and AI-supported teaching after using generative AI tools?

A thematic analysis of reflective journal and semi-structured interview data was conducted to gain an in-depth understanding of how teachers' perceptions are evolving. Reflective journals were used to document participants' ongoing experiences during intervention stages, while the interviews offered opportunities for more in-depth reflection and clarification of emergent ideas. Using an inductive coding method and after several rounds of refinement, four main themes were identified.

First, teachers saw AI increasingly as a catalyst for creative ideation. Instead of approaching AI as an assistant for creating existing content, participants described it as a source of inspiration that helped them push through creative constraints and explore new pedagogical opportunities. Numerous teachers shared that suggestions assisted with AI promotion inspired them to try new activity types and broaden their pedagogical toolkit. *"When things got tricky, ChatGPT came in and I was trying to come up with all sorts of possible creative things I had never come up with before."* (T3). This indicates that AI served as an intellectual accomplice, assisting teachers in refining their ideation, while human judgment and adaptation were retained.

Second, the people who answered the questions said they felt a difference in how they thought about planning their lessons in a new way that combines many things. The new way of planning lessons is used with the approach to their schoolwork. Because tools that use Artificial Intelligence helped teachers present information in both audio and text formats without pausing, they could make complex concepts simpler for students to understand. This was especially true for pictures, videos, and interactive materials created by Artificial Intelligence, which engaged students in class and allowed them to learn in different ways.

The teacher said, *"AI features from Canva have helped me turn ideas into something my students can actually see."*

These things show that AI does not just help people be more creative with their ideas. It also helps create interactive, dynamic teaching methods.

Despite these benefits, they also raised concerns about the authenticity and appropriateness of the AI-generated content. Some participants were skeptical of the naturalness, cultural relevance, and contextual suitability of AI outcomes for a variety of potential cultural contexts. Such language awkwardness or cultural misalignment was observed on some occasions in AI-generated materials, where even a teacher's training could translate into miscommunication in the classroom. *"Some generated AI sounded strange or didn't quite adhere to the cultural context."* (T6). These issues focus on evaluating the value of teacher assessments and draw attention to the view that AI is no substitute for the nuanced understanding that educators bring

to the classroom.

Lastly, teachers exhibited heightened awareness of self-reliance and the need to maintain pedagogical agency. Although AI was widely regarded as an excellent support tool, there were moments when participants realized how involved they had actually been in using it, and they stressed that even AI-produced work should still be adapted to their students' teaching needs. *“I learned I don't need to depend too much on AI, it's the context of my students that I still need to adjust everything to.”* (T2). This theme represents a developing equilibrium between using AI as a tool for efficiency and the need to uphold teacher autonomy in instructional decision-making.

Taken together, these findings imply that teachers' perceptions of AI shifted from a mere functional tool to AI as a collaborator in the process of making creative objects. Simultaneously, they fostered a critical and reflective perspective on the limitations of the practice, demanding contextually sensitive and pedagogically grounded use.

Table 2

Frequency of emergent themes from reflective journals (N=10)

Theme	Number of Mentions	Representative Quotes
AI as ideation partner	18	“AI helped me brainstorm lesson hooks.”
Multimodal enhancement	14	“I used AI images and dialogues to diversify input.”
Authenticity concerns	9	“Some AI examples lacked cultural fit.”
Dependence awareness	7	“I realized I shouldn't rely too much on AI.”

Overall, qualitative results supported the statistical analysis by revealing that teachers increased their creative practices and became more reflective in understanding the possibilities and drawbacks of generative AI in instructional design.

Summary of Findings

In conclusion, generative AI tools have significantly enhanced teachers' creativity, increasing the originality and versatility of their lesson designs. In this view, teachers see AI as a tool that enhances their ability to generate ideas and to encourage creativity in their lesson design. However, they also face a few issues, including concerns about authenticity and dependence on AI. Thus, both quantitative and qualitative findings highlight that AI has the potential to facilitate teachers' creativity, enabling them to improve the quality of their lessons. It is also important to note that these tools should be used critically, and the suitability of context-sensitivity within a pedagogical framework should not be neglected.

Discussion

AI as a creative co-designer

According to this study, generative AI can serve as a creative co-designer in teachers' lesson planning rather than a more passive role. The findings of this study, where creativity across

multiple dimensions improved significantly—notably originality and multimodal integration—demonstrate that AI can expand teachers’ ability to ideate, expand, and practice teaching in new and novel ways. From a theoretical standpoint, it is consistent with the notion of ideational fluency (Runco, 2014), which suggests the capacity to generate multiple ideas under time constraints. By exposing teachers to a wide array of prompts, representations, and teaching methods, AI appears to facilitate divergent thinking, a cornerstone of creativity. This echoes earlier arguments by Holmes et al. (2019), who suggest that AI systems can broaden users’ thinking horizons by introducing new perspectives. Crucially, the role of AI in this study is better captured through the framework of co-creativity (Long & Magerko, 2020), in which human and artificial intelligence agents work in synergy to achieve creative goals. Instead of replacing teachers’ know-how, AI facilitated idea generation, but teachers kept the reins on evaluation, revision, and pedagogical alignment. This dynamic demonstrates a transition from tools to collaborative cognition, in which creativity emerges from a balance between human judgment and machine-generated information. Whilst there is research supporting the efficacy gains associated with AI across studies, this study demonstrates an even more subtle role of AI - AI may allow teachers to shift cognitive effort away from routine steps in instruction to higher-order pedagogical decision-making. This result is supported by Yang et al. (2025) on the scaffolding role of AI in helping teachers build confidence and take creative risks -especially when challenged by time constraints or a lack of technical expertise. However, rather than treating AI almost exclusively as a time-saving tool, we see the technology as a facilitator of pedagogical innovation and a basis for experimenting with multimodal, interactive forms of instruction.

Balancing innovation and authenticity

While opening up creative possibilities, the adoption of generative AI also raises serious issues related to authenticity, cultural appropriateness, and contextual relevance. Teachers reported that some AI-generated content reflected Western-centric assumptions, including unfamiliar names, locations, and communicative norms that were not always relevant to Vietnamese learners. This finding points to an important limitation of generative AI in EFL practice: although AI can generate accurate language content, it does not automatically integrate local cultural nuances or pedagogical considerations into instruction. In this sense, teachers played a central mediating role in modifying AI outputs to be relevant to their surroundings. These adaptations have often taken the form of adapting cultural references, simplifying or localizing scenarios, or using language to incorporate students’ lived experiences. As such practices can be explained through the TPACK framework (Mishra & Koehler, 2006), this approach emphasizes the integration of technological, pedagogical, and content knowledge. Teachers’ skills in critically appraising and modifying AI-generated content reflect not only technological literacy but also pedagogical rationale and cultural fluency. Creativity, in this sense, is not just novelty but also appropriateness and contextual fit, consistent with socio-cultural theories of creativity (Sawyer, 2011). The results of this study also resonate with the concerns raised by Williamson and Eynon (2020), who are concerned about ‘automated creativity’, where ‘too much use of AI could result in less contextualized or superficial instructional design’. However, teachers in this study, rather than parroting AI output, engaged in active reinterpretation, indicating that professional agency is crucial in AI-empowered contexts.

Implications for educational leadership and teacher development

The implications of this analysis for leaders in the learning environment of educators and teachers are very significant.

First, school administrators must move past viewing AI as a specific technical tool and instead consider its importance in facilitating pedagogical innovation and innovative teaching. To that end, there is a need to establish institutional arrangements in which AI experimentation is supported by critical reflection on and encouragement of it.

Second, teacher training programs should include structured AI literacy opportunities for teachers to practice beyond simple use of the tools; they should also include opportunities for self-critical evaluation, awareness of the ethics of AI use, and the use of AI in various contexts. As this study shows, good use of AI also relies on teachers' ability to make sense of, adapt, and transform AI-generated outputs, rather than simply adopting them.

Professional development should also focus on teachers developing adaptive expertise, balancing efficiency and pedagogical intent. There should be scenario-based activities in which teachers can work through scenario-based training to develop ways to adapt AI content to individual learners' needs and cultural settings.

Lastly, institutions should consider embedding AI in collaborative professional learning communities for teachers from across professional roles, where they can dialogue about their practices, discuss barriers and challenges, and co-create tactics for real-world integration of AI in meaningful ways. There are also approaches that can be used to ensure AI as an adjunct in AI education, not a substitute for expertise from teachers, but to be part of, rather than a substitute, the creative and sensitive to the context and content of teaching and knowledge transfer when producing content.

Conclusion

The study indicates that generative AI could enhance teacher creativity in EFL lesson planning. Instead of simply reproducing AI-generated outputs, participants modified, improved, and contextualized them, resulting in more original, multimodal, and learner-centered instructional designs.

The results indicate that AI might serve as a creative co-designer, assisting idea generation while maintaining teachers' professional judgment and pedagogical agency. Moreover, the research adds to the expanding literature in AI and language education by applying AI and human co-creativity to teacher creativity and instructional design. It reinforces the idea that creativity in AI-based teaching requires novelty, contextual responsiveness, and pedagogical appropriateness.

However, a generalization should be made with caution in light of the study's relatively small sample size ($N = 10$) and its exploratory nature. More research with a broader, more diverse sample is needed across different educational settings to investigate the impact of AI on teacher creativity, professional identity, and classroom practice.

In conclusion, the results imply that the importance of generative AI for learning does not entail replacing the teacher; rather, it enables teachers to create more meaningful, creative, and context-specific learning experiences relevant to the classroom.

Acknowledgement

The authors would like to express their sincere appreciation to FPT University and Hoa Sen University for their academic guidance and institutional support. We would like to thank the teachers at Nguyen Van Cu High School, Hai Phong, for their passionate and helpful

participation in the professional development program. The authors are sincerely indebted to their colleagues for their constructive feedback and constant encouragement that helped to fulfil the completion of this study.

References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). <https://doi.org/10.1145/3442188.3445922>
- Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology. Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Craft, A. (2005). *Creativity in schools: Tensions and dilemmas*. Routledge.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., & Ahuja, M. (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, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Hoang, T. T. H., Dang, H. N., Truong, T. K. X., & Pham, N. B. Q. (2025). EFL teachers' perceptions of utilizing ChatGPT in designing lesson plans for IELTS reading skills. *International Journal of AI in Language Education*, 2(3), 19–39. <https://doi.org/10.54855/ijaile.25232>
- Hossain, M. K., & Al Younus, M. A. (2025). Teachers' perspectives on integrating ChatGPT into EFL writing instruction. *TESOL Communications*, 4(1), 41–60. <https://doi.org/10.58304/tc.20250103>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Jones, R. H., & Richards, J. C. (2015). Creativity and language teaching. In *Creativity in language teaching* (pp. 3–15). Routledge.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four-C model of creativity. *Review of General Psychology*, 13(1), 1–12. <https://doi.org/10.1037/a0013688>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376727>

- Luckin, R., Rodolph, J., Griner, M., & Tan, S. (2024). Exploring the future of learning and the relationship between human intelligence and AI: An interview with Professor Rose Luckin. *Journal of Knowledge Management and Education*, 7(1), 346–363.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Nguyen, H. T. M., & Burns, A. (2017). Teacher language proficiency and reform of English language education in Vietnam, 2008–2020. *Language, Education and Multilingualism*, 4, 93–107.
- Nguyen, L. A. D., & Le, T. T. P. (2025). Exploring the effects of an AI chatbot on emotional engagement in English speaking lessons: Insights from Call Annie. *International Journal of AI in Language Education*, 2(2), 79–99. <https://doi.org/10.54855/ijaile.25225>
- OpenAI. (2023). *GPT-4 technical report*. arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Pillai, A. K., Pereira, M., Roe, J., & Luong, L. K. (2022). Curriculum innovation in Vietnam: A study on teachers' attitudes using the dimensions of ACB (Affective, Cognitive and Behavioral) model. *Vietnam Journal of Education*, 6(1), 20–30. <https://doi.org/10.52296/vje.2022.141>
- Richards, J. C. (2013). *Creativity in language teaching: Perspectives from research and practice*. RELC Portfolio Series.
- Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice* (2nd ed.). Academic Press.
- Runco, M. A., & Jaeger, G. J. (2012). *The standard definition of creativity*. *Creativity Research Journal*, 24(1), 92–96. <https://doi.org/10.1080/10400419.2012.650092>
- Sawyer, R. K. (2011). *Explaining creativity: The science of human innovation* (2nd ed.). Oxford University Press.
- Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature*, 623(7987), 493–498. <https://doi.org/10.1038/s41586-023-06647-8>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Xie, X., & Fang, F. (2025). 'Walking out from the flat': Pre-service English language teachers' experiences of and needs for English as a lingua franca during study abroad. *Language and Education*, 39(3), 771–789. <https://doi.org/10.1080/09500782.2024.2378110>
- Yang, W., Li, H., & Lee, J. C.-K. (2025). Tailoring AI agents for early learning: The creative project approach. *Computers and Education: Artificial Intelligence*, 100473. <https://doi.org/10.1016/j.caeai.2025.100473>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). *Systematic review of research on artificial intelligence applications in higher education—Where are the*

educators? International Journal of Educational Technology in Higher Education, 16,
Article 39. <https://doi.org/10.1186/s41239-019-0171-0>

Biodata

Vu Thanh Loan is a lecturer of English at FPT University, Vietnam. She has over 12 years of teaching experience in English language education. Her research interests focus on technology-enhanced language learning, generative AI in ELT, teacher creativity, and professional development in computer-assisted language learning (CALL).

Nguyen Minh Tri is currently the Program Director of English Language at Hoa Sen University, Vietnam. He has extensive experience in English language teaching and teacher education. His research interests include language assessment, intercultural communication, academic writing, and innovations in English language pedagogy and curriculum development.

The Effects of ICTE Robot on Students' English-Speaking Skills: A Mixed-Methods Study on Non-English Majors at Nguyen Tat Thanh University

Ngoan Thi Pham^{1*}, Vu Phi Ho Pham²

¹ Nguyen Tat Thanh University, Vietnam

² Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam

*Corresponding author's email: ptngoan@ntt.edu.vn

 <https://orcid.org/0009-0009-4059-7706>

 <https://doi.org/10.54855/ijaile.26323>

® Copyright (c) 2026 Ngoan Thi Pham^{1*}, Vu Phi Ho Pham²

Received: 11/07/2026

Revision: 26/08/2026

Accepted: 31/08/2026

Online: 31/08/2026

ABSTRACT

Keywords: ICTE robot, AI Chatbots, speaking, feedback

Amid Vietnam's growing emphasis on English as a second language, educators continue exploring AI-powered tools to help learners speak accurately, naturally, and fluently. While prior studies have examined well-known chatbots such as ELSA, Google Assistant, and ChatGPT, few have investigated a customized chatbot within a mixed-methods design. This study addressed that gap by examining the impact of the ICTE robot, a custom-built AI chatbot, on the speaking performance and perceptions of 24 non-English-major students at Nguyen Tat Thanh University. Pre- and post-tests adapted from the PET speaking rubric assessed vocabulary, grammar, pronunciation, and discourse management, while semi-structured interviews with nine participants explored their experiences. Paired-samples t-tests revealed significant gains in vocabulary, pronunciation, and discourse management, though grammar showed no significant improvement. Interviews indicated that students valued the ICTE robot's voice-based and multimodal features, reporting greater confidence, stronger vocabulary retention, and improved pronunciation, while also noting technical issues such as speech-recognition errors and connectivity problems. These results suggest customized AI chatbots can meaningfully support EFL speaking development, highlighting the need for improved feedback accuracy and expanded grammar practice.

Introduction

Artificial Intelligence (AI) is undoubtedly assumed to impact all sectors, including education. Numerous educators are currently adopting AI-powered tools for speech cognition, such as ChatGPT and Gemini, as well as customized AI chatbots developed by scholars, which have enhanced second-language help and increased the effectiveness of teaching and learning. In particular, the emergence of AI chatbots such as the well-known ChatGPT, Gemini, and customized AI chatbots developed by scholars has boosted second-language acquisition. Martin (2025) asserted that the most pivotal feature of AI chatbots lies in Natural Language Processing

(NLP), which stimulates human-like communicative functions. Meanwhile, Halimah, Budianto, Hoesny, Cabuquin, and Ulfa (2025) identified 14 AI-based chatbots, including prominent ones such as Duolingo, Alice, and Elsa Speak, after screening 374 studies. Those previously mentioned chatbots are considered supportive tools for speaking skills.

Mastering oral communication is a multi-faceted challenge that extends far beyond a simple grasp of linguistic rules. Research by Heriansyah (2012) highlights that students confront both language-based hurdles, such as limited vocabulary and poor pronunciation, and psychological barriers like anxiety and shyness. Vietnam-based studies by Nguyen and Tran (2015) reveal that despite years of formal education, high school students often fail to achieve fluency due to performance pressure, knowledge gaps on topical issues, and fear of criticism. These observations are reflected in the author's experience at Nguyen Tat Thanh University, where even English majors struggle with in-depth discussions and instead offer brief responses. A primary factor is the inadequate opportunities to be immersed in an English-speaking environment.

In Vietnam, Decision 2371/QĐ-TTg was approved by the Prime Minister to make English a second language, posing a great challenge for both English instructors and learners. It means that English communicative competence, particularly English-speaking competence, should be highly regarded. Moreover, through observations of several general English classes, one of the two researchers realized that non-English majors were weak in pronunciation, vocabulary, grammar, and fluency. Regarding studies on the impact of AI chatbots on students' speaking skills, international and Vietnamese scholars employed various research designs, including experimental designs, surveys, and systematic reviews, but few conducted mixed-methods studies. Besides, they researched well-known AI chatbots rather than a customized chatbot like the ICTE robot.

Hence, an empirical study on the effectiveness of an AI-driven chatbot (a so-called ICTE robot) for oral improvement was conducted with the participation of non-English majors. The ICTE robot was invented to help students improve their vocabulary, grammar, pronunciation, and the four English skills. The competitive advantage of this chatbot is its voice feature, which helps students maximize their English learning.

The study was conducted to examine the impact of the ICTE robot on NTTU non-English-major students' vocabulary, grammar, pronunciation, and discourse management, as well as their perceptions of its use.

Literature Review

Definition of AI chatbots

An AI chatbot is a computer program that imitates human-user dialogues (Lexico Dictionaries, n.d, as cited in Adamopoulou & Moussiades, 2020). Altarif and Al Mubarak (2022) specifically described it as an AI tool that communicates with humans via messaging apps, mobile apps, and websites. AI chatbots have been defined by numerous scholars; however, in the realm of teaching and learning, chatbots are considered conversational agents that provide content via digital platforms (Bosch & Kruger, 2024).

Classifications of AI chatbots

According to Altarif and Al Mubarak (2022), AI chatbots are categorized as task-oriented, data-driven, and predictive. The former is merely associated with task fulfillment, whereas the latter seems more complex and collaborative. Task-oriented chatbots provide specific content based

on frequently asked questions, abide by rules, and use Natural Language Processing (NLP). Meanwhile, data-driven chatbots collect and analyze stored data to tailor content based on past performance. They can read texts, follow conversations, and give appropriate responses. The current study delved into task-oriented chatbots, originating from frequent questions to establish conversations with students.

Benefits and drawbacks of AI chatbots in education

AI chatbots are advantageous to both learners and educators; however, this study investigated how they benefit students. Meenakshi & Sharma, Bhavna (2026) noted that this type of technology facilitated personalized learning, enhanced accessibility, provided prompt feedback, and increased student engagement. Likewise, Gökçearsan, Tosun, and Erdemir (2024) agreed with the previous authors and added further positive aspects of AI chatbots, including increased motivation, enhanced language skills, and improved learning performance. Pitts, Marcus, and Motamedi (2025) presented fewer concepts and were geared towards time and cost efficiency, as well as creative support for students' further research. After reviewing them, all the scholars shared the view that, thanks to this type of technology, students can regulate their learning on their own, receive instant feedback without relying on teachers' scaffolding, and actively take on responsibility for learning.

Regarding the downsides of AI chatbots, they are perceived to hinder learners through misinformation, overreliance on technology, and restrictions on human interaction (Gökçearsan, Tosun, & Erdemir, 2024; Pitts, Marcus, & Motamedi, 2025; and Davar, Dewan, & Zhang, 2025). Pitts, Marcus, and Motamedi (2025) particularly raised their concern on academic integrity issue and students' loss of critical thinking skills while Gökçearsan, Tosun, and Erdemir (2024) distinctly presented major challenges which appear to surprisingly align with the current study, including "limitation of personalized feedback, failure to understand complex expressions, outdated responses, and difficulty getting used to technology" (p.25).

Definition of AI chatbots: Speaking and aspects of speaking skills

Definition of speaking

Linguists brought varied perspectives on the notion of speaking. Burns and Joyce (1997) identified it as the negotiation and establishment of meaning through the production and processing of information. Furthermore, depending on different circumstances, participants, and the speaker's intent, different meanings are conveyed. This is partly in line with Chaney and Burk's (1998) view of speaking as the use of verbal and non-verbal symbols to convey meaning. It seems that the previously mentioned scholars defined speaking broadly, leading to measurement issues in assessing it.

In contrast, Thornbury (2005) demonstrated a more systematic framework, regarding speaking as a combination of specific stages—conceptualization (think about discourse type, topic and purpose), formulation (choose discourse, syntax and vocabulary), and articulation (produce sounds)-facilitated by both linguistic knowledge (genre, discourse, pragmatic, grammar, vocabulary, phonology), and extra-linguistic knowledge topic and cultural knowledge, knowledge of the context and familiarity with other speakers. Torky (2006) broadened the concept by illuminating coherence, fluency, and appropriateness. In his view, sharpening speaking skills corresponds with sharpening linguistic, discourse, and pragmatic competencies. From other angles, Harmer (2001) emphasized the role of connected speech and expressive devices in conveying emotion and Kurum (2016) sorted the skill into three main components: mechanics (grammar, vocabulary and pronunciation), functions (the distinction between conveying information and maintaining interaction), and pragmatics (turn taking, rate of

speech, length of pauses of speakers, relative roles of participants). Ultimately, speaking is a complex skill, and defining it is troublesome. The operational definition of Kurum seems coherent, detailed, and measurable, which is used in the current study.

Aspects of speaking skills

There are two primary methodologies for classifying and assessing oral proficiency: the analytic and the synthetic approaches.

Referring to the analytic approach, Husnawati (2017) categorized oral performance into six key components.

Fluency: The capability of speaking at a natural speed with minimal hesitations, pauses, and use of time fillers.

Pronunciation: refers to how to produce utterances and includes pitch, intonation, individual sounds, sounds and spelling, and stress.

Grammar: refers to a system of structural rules which influence the composition of words, phrases, and clauses.

Vocabulary: relates to words, phrases, and expressions that show certain meanings. Firman and Haq (2012) clarified that vocabulary exists in active (words people use) and passive (words people comprehend) forms.

Interactive Communication: means the ability to interact with other people, such as asking and responding appropriately at the expected speed for task fulfillment. It reflects the promptness of learners' responses.

Appropriateness: means that, depending on setting, participant, purpose, and topic, people choose to communicate appropriately in terms of words and expressions.

Conversely, Koizumi (2005) analyzed only three elements: accuracy (the number of errors), fluency (temporal measures such as words per minute), and complexity (the varied and sophisticated use of vocabulary and grammatical structures).

The first classification provides more detail on learners' speaking competence and creates favorable conditions for raters. However, this one is prolonged and tedious for instructors of large classes. The second one is concise and efficient but short of meticulousness. The present study followed Husnawati's classification and examined vocabulary, grammar, pronunciation, and discourse management (equivalent to fluency), which are applicable to non-English majors. Due to the difficulties of online instruction and testing, interactive communication was not measured in this study.

The empirical studies on the impact of AI chatbots in EFL speaking classes and on students' perceptions of chatbot use.

Considering the promising potential of AI chatbots, numerous studies have been conducted both worldwide and nationwide. Aliakbari, Barzan, Allahveysi and Bakhtiarvand (2025) and Aliakbari, Barzan and Sayyadi (2025) concurred in adopting an experimental design, but the second research article had more specific findings that intermediate-level students progressed slightly in pronunciation and had remarkable improvement in fluency, vocabulary, and grammar after 12 weeks of practice with LinguaBot. In addition to pre- and post-tests, other researchers used questionnaires and semi-structured interviews to collect comprehensive data. Zhang (2026) allowed students to experiment with CallAnnie, Google Assistant, and Alexa to evaluate Chinese students' oral proficiency and afterward conducted semi-structured interviews

to gain their perceptions on the use of AI chatbots and willingness to communicate. Having analyzed the ANOVA data and interview responses, the author found that students' oral proficiency improved and that they appreciated "AI chatbots' accessibility, authenticity, and multimodal features, which helped reduce anxiety and boost confidence" (p. 1). Similarly, Grab (2025) explored four AI chatbots, namely Talkpal AI, Gliglish AI, Langotalk, and ChatGPT (voice mode), and administered post-study questionnaires and interviews. The findings revealed that those tools positively impacted students' speaking skills and increased their confidence; personalized feedback and autonomous learning were factors they highlighted after the intervention.

Using another research method- a systematic review, Safitri, Hidayati, and Ciptaningrum (2025); López-Molines (2025) and Rachmadani, Musyafa'ah, Januanto, and Fitriyah (2026) concurred in concluding that AI chatbots were effective tools in speaking instruction. Regarding students' perspectives on the chatbots, López-Molines's findings appeared to be the most meticulous. The author affirmed that thanks to chatbots, students boosted their confidence and performed better in language learning at their own pace. Conversely, they had shortcomings, including incoherent responses, poor speech recognition, and poor sound quality. Safitri, Hidayati, and Ciptaningrum (2025) highlighted the advantages, including increased learner participation and opportunities for practice, and the disadvantages, including response inadequacies. Rachmadani, Musyafa'ah, Januanto, and Fitriyah failed to discuss the drawbacks of the chatbots. In terms of methodology, López-Molines and Safitri, Hidayati, and Ciptaningrum followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, while the remaining authors used the Search, Appraisal, Synthesis, and Analysis (SALSA) framework.

In the Vietnamese context, mixed methods were predominantly employed among the researchers. Tran (2025) engaged 50 private English majors in completing survey forms and, for some, in interviews after learning English with ChatGPT, Elsa Speak, and Duolingo for 6 weeks. Besides the previously mentioned findings, such as automatic mistake identification and correction, concerns about the cost of chatbots are emerging. Le (2025) adopted the identical research method to Tran (2025). Bui, Ngo, and Mai (2025) conducted a large-scale study on the effects of ChatGPT and Character AI on the argumentation skills of English-major students, while Le and Ngo (2025) investigated the effects of Yoodli, Orai, and other AI tools on English-major students' presentation skills. The similarities between the two studies include the use of both quantitative and qualitative methods and the identification of both the pros and cons of chatbots. Particularly referring to the speaking research topic, Duong and Suppasetseree (2024) conducted a quasi-experimental investigation, combined with the use of questionnaires and interviews, to delve into the effectiveness of the Andy English bot in improving oral capabilities and students' attitudes toward this chatbot. The study found that students' accuracy and fluency were improved, and that the comfort and flexibility in practicing with the chatbot were valued.

After reviewing both domestic and overseas studies, it was found that all scholars investigated well-known AI chatbots, and few conducted mixed-methods studies. The present study focused on a customized chatbot developed by one of the two authors, aimed at improving students' vocabulary, grammar, pronunciation, and discourse management, with participation from the participation of non-English majors at a private university.

Research questions

To address those objectives, two research questions were raised:

1. To what extent does the ICTE robot improve NTTU non-English majors' speaking performance, including vocabulary, grammar, pronunciation, and discourse management?
2. What is the perception of NTTU non-English majored students on the adoption of the ICTE robot in speaking class?

Methods

Participants and setting

Twenty-four non-English major students voluntarily participated in this study after obtaining official approval from the Faculty of Foreign Languages. One of the two researchers who directly taught those students obtained the students' consent to participate in the study, including recording the pre- and post-tests, storing screenshots of practice with the ICTE robot, and recording the interviews. Seventeen males and seven females were aged 20-22 and mostly studied automotive engineering technology and information technology engineering. All of them were taught by one of the two researchers in the AVCD4 class. Their English proficiency was determined by the pre-test score. Convenience sampling was chosen to reduce time and expenditure (Golzar et al., 2022). The study's objectives and the students' assignments were clearly explained to them. Every participant was instructed in the use of the ICTE robot and had a pre-intermediate level of English proficiency. After finishing each unit, the students were encouraged to practice using the ICTE robot to consolidate their knowledge and improve their English skills. The study was conducted over 17 sessions, comprising 8 online and 9 offline sessions, in compliance with the University's direction during semester 2 of the 2025-2026 academic year. The participants were taught with the Personal Best B1 coursebook (units 7-12).

Research design and instruments

A mixed- methods study was adopted, combining quantitative and qualitative analyses. According to Creswell (2012), this research method provides comprehensive insights into a phenomenon.

The researchers used pre- and post-tests to assess the ICTE robot's effectiveness in improving students' speaking skills, and semi-structured interviews to explore their perceptions of using it. The pre-test and post-tests were drawn from the sample PET speaking test, using the same format and content to ensure validity. Taking into account the students' English proficiency and time constraints, only Parts 1 and 2 were administered. In Part 1, students were required to answer general questions, and in Part 2, they completed a picture description. The tests were marked by two raters to increase the validity of the data. Both had extensive experience in teaching the speaking module for non-English majors. They also previously administered PET speaking tests. The rubric was adapted from the PET speaking rating rubric to ease the assessment. In the PET speaking rubric, four dimensions of speaking (vocabulary, grammar, pronunciation, discourse management, and interactive communication) are assessed, and students' performance band scores range from 0-5. The raters focused on vocabulary, grammar, pronunciation, and discourse management, using a 0-5 scale for each criterion. Considering students' proficiency and virtual learning mode, the interactive communication criterion was not marked.

The Kappa coefficient indicating the agreement level between the two rater is presented in the table below.

Table 1

The inter-rater reliability

		Value	Asymtotic standard error	Approximate T	Approximate significance
Measure of agreement	Kappa	.03	.13	.28	.78

As described, the Kappa value is 0.03 (<0.6), indicating that there is no significant agreement between the two raters. The final score was determined by the first rater by calculating the average score or marking based on either rater, after carefully reviewing the recorded speaking performance audio.

The ICTE robot was invented by one of the two authors, who used ChatGPT to code. There are eight features, including conversation, vocabulary, grammar, discussion, reading, listening, problem-solving, and progress. The students were encouraged to exploit all the functions to maximize their learning outcomes. The vocabulary, grammar, listening, and reading exercises are standard but the additional tasks in the ICTE robot is conversation (students interact with the robot and are commented on responses); discussion (students use the learned vocabulary and structures to practice writing), problem-solving (students deal with real-world situations), and progress (check the number of completed activities, average score and the number of words used). The notable feature is that students voice chat with the robot like a native speaker rather than merely doing tedious exercises, and they can use the voice function to learn every lesson. The link to access the ICTE robot is <https://i-cte.org/robot>. To help readers understand those features of the ICTE robot, the screenshots in Unit 7 (City living) are as follows.

Figure 1

Interface of ICTE robot

The figure displays two screenshots of the ICTE robot interface, specifically Unit 7 — City Living & The Daily Commute. The interface features a green navigation bar with tabs for Conversation, Vocabulary, Grammar, Discussion, Reading, Listening, and Problem-solving. The main content area is divided into sections for practice exercises.

Top Screenshot: Vocabulary Matching + Use It

Unit 7 — City Living & The Daily Commute
Practice present perfect with *yet / already*, city features, commuting vocabulary, and listening for facts & figures.

Mic: Off
Start Voice Stop
Voice (Narrator)
Google US English (en-US)
Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

2) Vocabulary Matching + Use It
Read instructions Reset

Match the words to the meanings, then complete the sentences using the same vocabulary. (City features list :contentReference[oaicite:4](index=4); transportation list :contentReference[oaicite:5](index=5))

A) Matching

1) bike lane
Choose meaning...

2) traffic jam
Choose meaning...

3) platform
Choose meaning...

4) commuter
Choose meaning...

B) Use the words (Fill in the blanks)

1) I always walk on the ___ when the road is busy.
Choose...

2) At the ___, I wait for the train and then get on.
Choose...

3) During ___, the streets are crowded and commuters get delayed.
Choose...

Bottom Screenshot: Grammar — Present perfect with yet / already

Unit 7 — City Living & The Daily Commute
Practice present perfect with *yet / already*, city features, commuting vocabulary, and listening for facts & figures.

Mic: Off
Start Voice Stop
Voice (Narrator)
Google US English (en-US)
Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

3) Grammar — Present perfect with *yet / already*
Read instructions Reset

Rule reminder: Use **already** for something earlier than expected; use **yet** in questions/negatives for something expected before now (present perfect focus :contentReference[oaicite:6](index=6)).

1) Have you found a job ___ ?
Choose...

2) I've ___ been to the theater twice.
Choose...

3) She hasn't arrived ___ .
Choose...

4) I've ___ paid the rent.
Choose...

January 28, 2026 admin

Conversation Vocabulary Grammar **Discussion** Reading Listening Problem-solving

Unit 7 — City Living & The Daily Commute

Practice **present perfect** with *yet / already*, city features, commuting vocabulary, and listening for **facts & figures**.

Mic: Off

Start Voice **Stop**

Voice (Narrator)

Google US English (en-US)

Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

4) Discussion (Voice or typing)

Read instructions **Next question**

Answer with reasons. Try to include at least **2 vocabulary items** and at least **one** present perfect sentence using *yet* or *already*.

Click Next question to begin.

Write your answer here...

Check my language **Read question**

← IELTS – Listening – Test 6 ICTE Tools & Conferences →

Conversation Vocabulary Grammar Discussion **Reading** Listening Problem-solving

Unit 7 — City Living & The Daily Commute

Practice **present perfect** with *yet / already*, city features, commuting vocabulary, and listening for **facts & figures**.

Mic: Off

Start Voice **Stop**

Voice (Narrator)

Google US English (en-US)

Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

5) Reading Comprehension — “First impressions of city life”

Read instructions **Read the text**

Read the text, then answer the questions. (Source text about Alex & Megan :contentReference[oaicite:7](index=7))

San Francisco — first impressions

- There are so many tech companies here; it's really exciting. I've joined a networking group to meet people from the industry, and I've already been to a couple of events. Have I found a job yet? No ... but I'm sure I will soon!
- In my apartment building, there are people of all nationalities. I haven't met all my neighbors yet, but I love living in a city where everyone is different.
- Whatever you're interested in, there's always something going on. I've already been to the theater twice and I've seen a baseball game, and I only arrived two weeks ago!
- San Francisco is expensive and you need a good salary to live here. I've rented a room in a shared house, but after I've paid the rent, it doesn't leave me with much money for other things.
- The summers wouldn't be as hot as back home. It hasn't been too cold yet, but it can get really foggy, especially when you're near the ocean.

ConversationVocabularyGrammarDiscussionReadingListeningProblem-solving

Unit 7 — City Living & The Daily Commute

Practice **present perfect with yet / already**, city features, commuting vocabulary, and listening for **facts & figures**.

Mic: Off

Start VoiceStop

Voice (Narrator)

Google US English (en-US)

Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

6) Listening Comprehension — Commuting around the world

Read instructions

Play listening (TTS)Stop

Click **Play listening**. Listen for **facts and figures**, then complete the gaps. Target answers include: *40 minutes, two hours, 7:30, eight million, push, white, 40 million, cars, argued, big* (answer set :contentReference[oaicite:9](index=9)).

Listening focus: identify numbers/times and key nouns (e.g., *platform, passengers, parking space, public transportation*). (Skill + vocab focus :contentReference[oaicite:10](index=10))

Word bank: 40 minutespush7:30eight millionarguedtwo hoursbigcarswhite40 million

1) The global average commuting time is ____.

Choose...

2) In Bangkok, the average commuting time is ____.

Choose...

VocabularyGrammarDiscussionReadingListeningProblem-solvingProgress

Unit 7 — City Living & The Daily Commute

Practice **present perfect with yet / already**, city features, commuting vocabulary, and listening for **facts & figures**.

Mic: Off

Start VoiceStop

Voice (Narrator)

Google US English (en-US)

Tip: If you don't see voices yet, click anywhere once and wait 2–3 seconds.

7) Problem-solving (Real-life city scenarios)

Read instructionsReset

Choose the best solution. Then write one sentence using present perfect with *yet* or *already*.

1) Scenario 1: You start out late and there's a traffic jam at the intersection near the bridge. You must arrive on time. What's the best plan?

☐ Drive faster through the crosswalk to save time.

☐ Take public transportation and walk from the station platform to your office.

☐ Park in a bike lane so you can get out quickly.

Write ONE sentence using present perfect with *yet/already* about what you have done to fix the problem.

Write your sentence here...

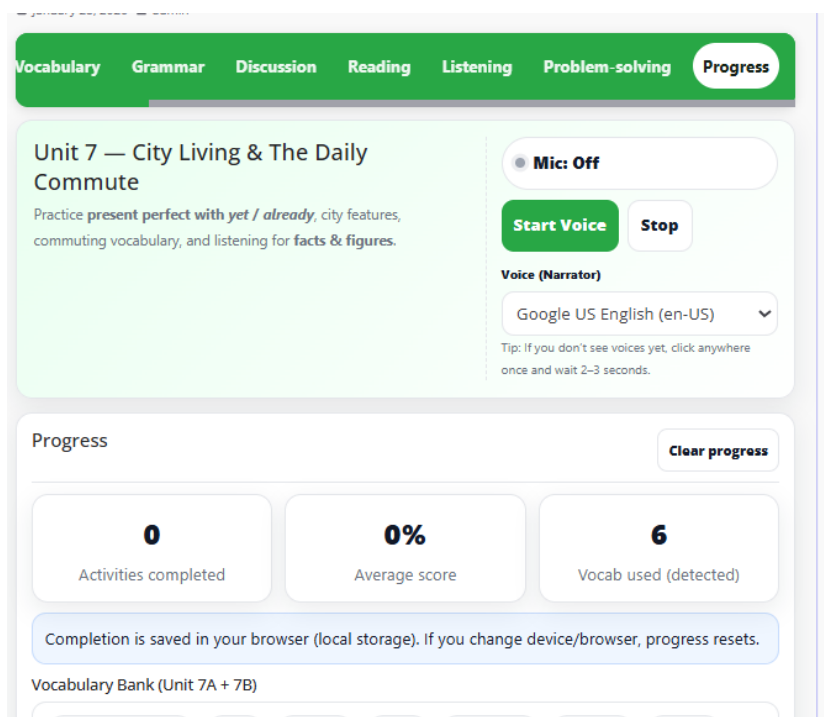
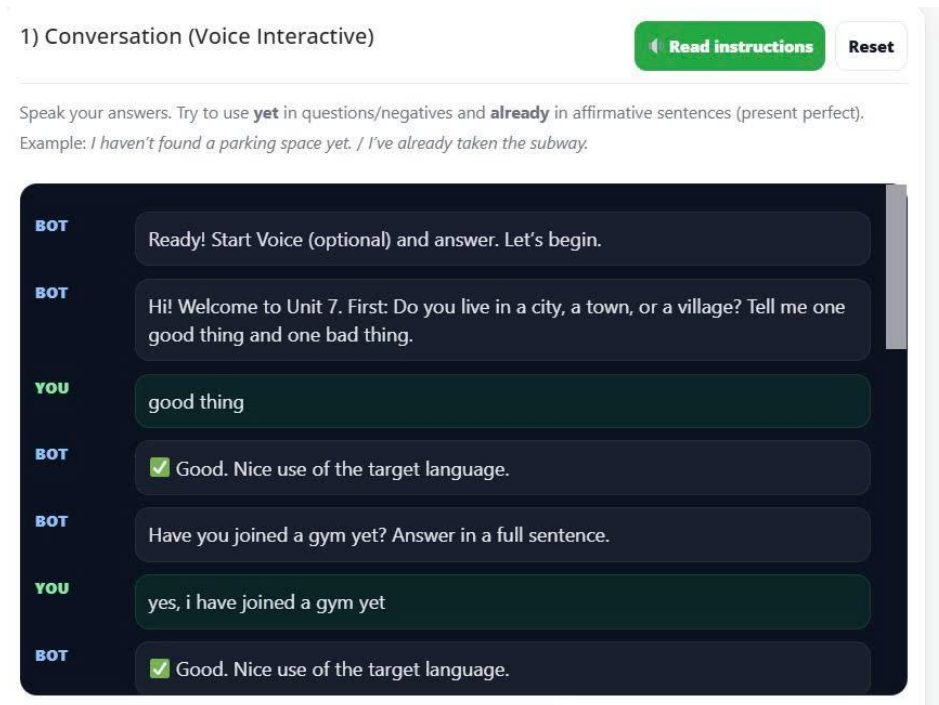


Figure 2 shows evidence of one student's practice with the ICTE robot, in which the student practiced the Conversation section in Unit 7 and received instant feedback from the robot.

Figure 2

Students practiced with the ICTE robot



Eleven interview questions were translated into Vietnamese to facilitate participants' responses.

When do you usually practice speaking English with the chatbot? How long do you practice?

Do you find that practicing with the chatbot helps you speak English more confidently?

Do you find learning English with a chatbot similar to speaking with a native speaker?

Which learning method do you prefer?

Do you find that learning through the chatbot helps you improve your vocabulary, pronunciation, and reflexes?

What aspects of the chatbot do you like?

What difficulties have you encountered when learning with the chatbot outside of class?

Do you have any suggestions for the chatbot?

Do you think the chatbot is a powerful tool to support learning to speak English?

In your experience, what benefits did practicing English speaking with the chatbot during the course bring to you?

Do you think we should continue incorporating this chatbot to supplement the English-speaking practice in the current curriculum?

At first, the researchers intended to choose ten participants for the interview process, but one of them declined. The selection of nine interviewees was based on their performance and engagement in the lesson. All low-, medium-, and high-engagement learners were included. To assess the saturation of the interview codes, the researcher interviewed the first 3 participants, coded the interviews, and realized there was insufficient information about their learning experience with each function of the ICTE robot. Subsequently, the researcher conducted interviews with the next 3 participants, coded, and found that they did not share much about the benefits and challenges of using the ICTE robot. Finally, the remaining participants were interviewed to ensure sufficient data were collected.

Data collection procedure

This procedure underwent two phases: intervention and interview administration. Initially, pre-tests were administered virtually during the first three sessions, and all audio recordings were obtained with students' consent. After completing one unit, the participants practiced English with the ICTE robot to improve English proficiency, focusing on speaking skills. They could freely choose their preferred section in the ICTE robot to practice and learn at their convenience. After that, they took screenshots and sent those images to the teacher as evidence. The detailed learning activities of six units were described below.

Session	Content	Learning activity
1	Unit 7A+B	Have a group discussion about life in the city
2	Unit 7C+D	Ask questions using the structure Have/has + V3
3	Unit 8A+B	Ask questions about the food topic
4	Unit 8C+D	Make a dialogue using the structure Would you like to V
5	Unit 9A+B	Make sentences using the structure Used to V
6	Unit 9C+9D	Have a group discussion about online shopping
7	Unit 10A+10B	Make a short conversation about Sports
8	Unit 10C+10D	Make sentences using the structure reported speech
9	Unit 11A+11B	Ask questions using the structure Gerund/ Infinitive
10	Unit 11C+11D	Make sentences to describe the materials
11	Unit 12A+12B	Make sentences using the relative pronoun
12	Unit 12C+12D	Have a group discussion about the First date

This intervention lasted 12 sessions and was followed by the post-test. Finally, based on the evidence of practice submitted to the teacher, 9 participants (5 females and 4 males) were purposefully selected for the recorded interview process. Females were coded as F and males were coded as M in the qualitative data analysis. Each participant was interviewed separately in 3 sessions. The researcher conducted the interview in Vietnamese to create favorable conditions for them to share their thoughts on the use of the ICTE robot. There was only one coder in the qualitative analysis.

One of the researchers conducted the interviews and transcribed the interview audio for analysis. The transcript was not returned to the participants, and the researchers translated only some of the sentences included in the qualitative analysis into English. An expert in English Language Teaching was selected to check the translation. The analysis of qualitative data was therefore easier. The qualitative methodology of this study was adapted from the fundamental steps of thematic analysis by Almedia (2026).

Step 1: in-depth reading of AI chatbot studies on improving students' English, highlighting studies that include interview questions, such as Zhang (2026). Some themes were defined, such as feelings about using AI chatbots outside the classroom, the impacts of AI chatbots on communication in English, the difference between communicating with AI chatbots and communicating with humans, and the challenges of using AI chatbots outside the classroom

Step 2: Adding more themes such as frequency of practice with the ICTE robot, the effects of the ICTE robot on vocabulary, grammar, pronunciation, and discourse management, perceptions of the multimodal features of the ICTE robot, benefits and drawbacks of using the ICTE robot, and the recommendations)

Step 3: Interviewing session 1 and initial coding process (after interviewing the first 3 students, the researcher transcribed and put the codes in each theme)

Step 4: Interviewing session 2 and coding

Step 5: Interviewing session 3 and coding

Step 6: After transcribing the interview audio, the researchers highlighted key points and created a table listing each interviewee and the interview themes.

Results/Findings

Research question 1: To what extent does the ICTE robot improve NTTU non-English majors' speaking performance, including vocabulary, grammar, pronunciation, and discourse management?

Pre-test

Table 2 presents descriptive statistics for pre-test scores in vocabulary, grammar, pronunciation, and discourse management.

Table 2

Descriptive statistics of pre-test scores

Variables	N	Minimum	Maximum	Mean	Std. Deviation
Vocabulary pre-test	24	.5	3.0	1.77	.63
Grammar	24	.0	2.5	1.40	.71
Pronunciation	24	.0	3.0	1.52	.77
Discourse management	24	.0	3.0	1.13	.85
Valid N (listwise)	24				

As described, before learning with the ICTE robot, the mean vocabulary score was the highest at 1.77 (SD=0.63), followed by pronunciation, grammar, and discourse management, respectively. To be specific, prior to treatment, the mean pronunciation score (M=1.52, SD=0.77) was slightly lower than the mean vocabulary score, and the mean discourse management score (M=1.13, SD=0.85) was a bit lower than the mean grammar score (M=1.40, SD=0.71). It means that before being supported by the ICTE robot, the students' ability to use words in speaking was the best, and their fluency was the worst.

Post-test

Table 3 presents the descriptive data of vocabulary, grammar, pronunciation, and discourse management after practicing with the ICTE robot.

Table 3

Descriptive Statistics of post-test scores

	N	Minimum	Maximum	Mean	Std. Deviation
posttest vocabulary	24	1.0	3.5	2.19	.62
posttest grammar	24	.0	3.5	1.54	.90
posttest pronunciation	24	1.0	3.0	2.06	.63
post test discourse management	24	.0	4.0	1.92	1.06
Valid N (listwise)	24				

The table shows that after learning with the ICTE robot, the students continued to perform best in using lexical resources (M=2.19, SD=0.62) and worst in applied grammatical structures (M=1.54, SD=0.90). The mean of pronunciation score (M=2.06, SD=0.63) and discourse management (M=1.92, SD=1.06) were nearly as high as that of vocabulary, indicating that the proficiency in pronunciation and fluency was approximately equal to vocabulary.

Paired samples T-test

The comparison of vocabulary, grammar, pronunciation, and discourse management scores before and after learning with the ICTE robot was shown in Table 4.

Table 4.

Paired samples T-test

	Cohen's d	Mean	Std. Deviation	t	df	95% Confidence Interval of the Difference		Sig. (2-tailed)
						Lower	Upper	
Pre-test vocabulary & post-test vocabulary	.67	-.42	.67	-3.05	23	-.70	-.14	.01
Pre-test grammar & post-test grammar	.85	-.15	.85	-.84	23	-.51	.21	.41
Pre-test pronunciation & post-test pronunciation	.78	-.54	.78	-3.41	23	-.87	-.21	.00
Pre-test discourse management & post-test discourse management	1.25	-.79	1.25	-3.10	23	-1.32	-.26	.01

The table depicts that the p-values of most elements, including vocabulary, pronunciation, and discourse management, were lower than 0.05, meaning that there was a significant difference in the results of those speaking aspects between before and after the intervention. Conversely, there was no significant difference in grammar before and after integrating the ICTE robot in the learning program ($p=0.41$). It can be inferred that the ICTE robot was an effective tool in improving students' lexical resources, pronunciation, and fluency. Specifically, regarding the mean difference and Cohen's d indicating the effect size, the students had a considerable improvement in discourse management ($M= -0.79$, $\text{Cohen's } d = 1.25$) while they had a slight improvement in pronunciation ($M=-0.54$, $\text{Cohen's } d = 0.78$), and vocabulary ($M= -.42$, $\text{Cohen's } d = 0.67$), respectively.

Research question 2: What is the perception of NTTU non-English majored students on the adoption of the ICTE robot in the speaking class?

From the participants' perspectives on the issues in using the ICTE robot to improve their speaking skills, the researchers identified five dominant themes in the responses of the interviewed students.

To what extent does the ICTE robot help students boost their vocabulary, grammar, pronunciation, and discourse management?

According to the participants, the ICTE robot significantly improved pronunciation, prolonged word retention, and slightly improved syntactic structures.

After practicing with the ICTE robot, I feel that I pronounced better. In terms of vocabulary, I know more new and interesting words. I not only understand them but also know how to use them in context" (F1).

I didn't improve a lot, but I feel the extent of the improvement. About vocabulary, I improved the most. When I remember the words, I can read and write more quickly. About grammar, it just improved a bit (M1).

After learning with the ICTE robot, I can remember words better. And my ability to pronounce words has greatly improved because I normally miss final sounds, but when practicing with the robot, I tried to pronounce ending sound correctly (F3).

In short, thanks to the ICTE robot, the students improved their vocabulary and pronunciation dramatically but only marginally improved their grammar.

What do students think about the multimodal features of the ICTE robot?

All of them appreciated the ICTE robot's voice feature, which helps them record their speech and interact with the chatbot. It is supposed that it is an extraordinary and emergent finding.

First of all, some of them asserted that vocabulary tasks were plentiful and exciting.

There are many types of vocabulary exercises and there is much interesting content (F2)

Secondly, they complimented the design of the ICTE robot which eases comprehension.

The interface of the grammar section is appealing, which helps me understand the structures and learn English easily (M2).

Thirdly, thanks to the voicing and recording features, they can practice pronunciation and listening skills in all sections of the chatbot.

In the vocabulary section, students can practice placing stress on syllables and words (M1).

In the conversation section, I can talk with the robot. The sound is clear and loud enough. I can improve listening skills, and the robot corrects my mistakes (F5).

In short, all the participants approved of the voicing feature and multitasking functions of the ICTE robot.

Which sections of the ICTE robot were used frequently?

Most of them selected practice sections to suit their preference and needs. Two of them preferred the Conversation section.

I like practicing communication. Through the meaning, I can find suitable words to interact with the ICTE robot (F1).

In the conversation section, I can communicate with the ICTE robot, and practice speaking and pronunciation. The robot can correct my mistakes (F4).

Most of them chose to learn on vocabulary, reading and grammar sections.

I am not good at vocabulary, listening and speaking. So, I focus on those sections (M1).

In the vocabulary section, I can review words. In the reading section, I can recall the lesson and vocabulary to remember them longer. I can record my voice and talk with the ICTE robot (F3).

In the grammar section, there are examples so I can remember the structures easily (F5).

In brief, vocabulary, reading, and grammar were the students' preferred section in the ICTE robot.

What are the benefits and drawbacks of the ICTE robot?

All the interviewees asserted that their confidence has been boosted by the ICTE robot.

The chatbot is not human, so I am not scared of making mistakes (M3).

After learning with the robot, I am a bit more confident (F3).

The other plus points of the ICTE robot are its flexibility, prompt feedback, and ability to detect and correct errors.

I can learn with the ICTE robot anytime, anywhere. And when I make mistakes, it corrects them right way” (M1).

After learning with the ICTE robot, I know how to pronounce correctly and give correct responses (F2).

There are some new findings, such as the chatbot's reasonable speed to facilitate listening comprehension, an AI assistant to handle inquiries, maximizing learning, and increased excitement through diverse tasks.

When you combine the ICTE robot in- class lessons, I feel I can understand the lesson better. I can do various tasks: vocabulary, reading, listening...(M2)

The speed of the ICTE robot is slow, so I can listen easily (F2)

Because there are many tasks, learning is not boring. I am more eager to learn English (F3).

The chatbot helps me improve listening comprehension and answer the questions correctly. Besides, when I have questions, I can ask the AI assistant (F5).

However, one commented that the interface is not neat enough, resulting in confusion.

Regarding the chatbot interface, because too many things appear at once, I feel confused, and sometimes it is difficult to find the section to practice (F1).

Most of the interviewees exhibited negative attitudes toward technical problems such as slow or no responses, failure to identify students' voices, microphone errors, and voice-off failures.

Sometimes I can't turn off the mic, or the robot can't hear my responses, or it responds incorrectly (F4).

Sometimes the robot says non-stop, so I don't understand (F5).

Some of them were not content with the chatbot's incorrect feedback and low Internet access.

Sometimes I pronounce words correctly, but the robot says it is wrong. Or I say this word, but the robot identified that word. I feel that it gave wrong feedback (M1).

What I don't like about the chatbot is that sometimes it lagged, or the mic was poor quality (M4)

Generally, they agreed that the ICTE robot is useful for their English learning and should be applied in the future.

What are the recommendations to improve the efficacy of the ICTE robot?

They predominantly argued that the technical problems mentioned should be addressed.

I think the conversation section should be upgraded. I mean the voice function should be upgraded (M2).

I think you should adjust the speed of the chatbot (F5).

One recommendation is that the researcher add more levels of difficulty to the exercises and design additional exercises beyond the coursebook so that students at mixed levels can practice.

I think that the teacher created the robot based on the coursebook content. But in our class, many students are not good at English. So, when they practice with the chatbot, they can feel that it is difficult. So, the teacher can design more exercises at a lower level of difficulty. All of us can learn (F3).

In summary, most students affirmed that addressing technical issues would enhance the ICTE robot's effectiveness.

Discussion

The finding for RQ1 is partly similar to that of Aliakbari, Barzan and Sayyadi (2025). This can be explained by the fact that, in this study, students' pronunciation slightly improved thanks to the ICTE robot's voice feature, which enabled them to record their speech and correct pronunciation errors. However, the repeated speech-production opportunities were limited, and speech-recognition feedback was sometimes incorrect, leading to minimal improvement. The lecturers can exploit the ICTE robot to its full potential to help non-English -major students improve their pronunciation, which is considered the weakest aspect. They can not only progress in pronunciation but also in word and sentence stress. The difference was that the present study showed no improvement in grammar, whereas the previous study showed moderate progress. Low proficiency in grammar may relate to limited practice intensity or poor task design. The researchers could make some innovations to this section of the ICTE robot.

The first result of RQ2 is somewhat at variance with that of Aliakbari, Barzan and Sayyadi (2025). The interviews indicated significant gains in vocabulary. Vocabulary gains may relate to repeated retrieval and contextual reuse. In other words, the students learn vocabulary through repeated tasks and know how to use the words in real contexts.

Regarding the benefits of the ICTE robot, the responses are supported by Zhang (2026), Grab (2025), and Tran (2025), who found that using the AI chatbot is beneficial for boosting confidence, flexibility, instant feedback, error detection, and correction. The similarity in findings can be explained by the fact that these are shared features among various AI chatbots. For instance, it takes at least one day to receive the teacher's feedback, whereas the chatbot can detect users' voices and errors in just a few seconds and provide feedback at once.

With reference to the downsides of chatbots, they concurred that the AI chatbot is somewhat disadvantageous to them due to technical problems (slow or no responses, no identification of students' voices, microphone errors, and voice-off failure), poor Internet access, and incorrect feedback. López-Molines (2025) and Safitri, Hidayati, and Ciptaningrum (2025) reported results similar to those of the current study, as machine-based interactions cannot avoid technical problems, regardless of the devices used. It is advisable to minimize those issues to create more favorable conditions for students to learn English with the ICTE robot.

Conclusion

To examine how the ICTE robot affected students' speaking skills and how chatbot use was perceived, a combination of quantitative and qualitative instruments was used. The quantitative findings revealed that the ICTE robot improved students' vocabulary and pronunciation, and

the qualitative findings showed that this digital tool helps them learn and use more new words, understand structures better, and pronounce words correctly. However, fluency practice should be carefully planned and systematically conducted to help students improve their speaking comprehensively. The advantages of the ICTE robot seemed to outweigh the disadvantages. They proposed that the technical problems be addressed, the ICTE robot's speed be adjusted, and more graded exercises be created. The study had some unavoidable limitations, such as a small sample size, virtual testing before practicing with the ICTE robot, and partial test administration (only two of four parts were tested). One of the major limitations is the use of a single-group experiment, which struggles to control other variables that affect students' speaking improvement. Despite those constraints, this research makes a contribution to the field of AI applications in English teaching.

Acknowledgments

The authors would like to take this opportunity to express their deepest gratitude to Nguyen Tat Thanh University, particularly to the study participants, for their efficient support. Without their support, this research could not be accomplished.

References

- Adamopoulou, E., Moussiades, L. (2020). An overview of chatbot technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds). *IFIP Advances in Information and Communication Technology*, 584, 373-383. https://doi.org/10.1007/978-3-030-49186-4_31
- Aliakbari, M., Barzan, P., Allahveysi, S. P., & Bakhtiarvand, M. (2025). AI-powered chatbots for English conversation practice: Impact on speaking proficiency. *Technology Assisted Language Education*, 3(3), 45-62. <http://doi.org/10.22126/tale.2025.12621.1138>
- Aliakbari, M., Barzan, P., & Sayyadi, M. (2025). Exploring the impact of AI chatbots on EFL learners' conversational proficiency. *Journal of Interdisciplinary Research in English Language Communication*, 1(2), 66-80. <https://doi.org/10.30470/irelc.2025.2058800.1022>
- Almedia, F. (2026). Thematic analysis in qualitative research: Steps, benefits, and challenges. *Journal of Action Qualitative & Mixed Methods Research*, 5(1), 25-38. <https://doi.org/10.5281/zenodo.18241378>
- Altarif, B., Al Mubarak, M. (2022). Artificial Intelligence: Chatbot—The New Generation of Communication. In: Hamdan, A., Harraf, A., Arora, P., Alareeni, B., Khamis Hamdan, R. (eds) *Future of Organizations and Work After the 4th Industrial Revolution*. Studies in Computational Intelligence, vol 1037. Springer, Cham. https://doi.org/10.1007/978-3-030-99000-8_12
- Bosch, C. & Kruger, D. (2024). AI chatbots as open educational resources: Enhancing student agency and self-directed learning. *Italian Journal of Educational Technology*, 32(1), 53-68. <http://doi.org/10.17471/2499-4324/1326>
- Burns, A. and Joyce, H. (1997). *Focus on speaking*. Sydney: National Centre for English Language Teaching and Research.

- Bui, T. T. G., Ngo, N. V. A., & Mai, D. H. (2025). Applying AI tools to enhance argumentation skills in debate activities for EFL Students. *International Journal of AI in Language Education*, 2(3), 83-100. <https://doi.org/10.54855/ijaile.25236>
- Cardona, M. A., Rodríguez, R. J. & Ishmael, K. (2023). *Artificial Intelligence and future of teaching and learning: Insights and recommendations*. U.S. Department of Education, Office of Educational Technology.
- Chaney, A. L. & Burk, T. L. (1998). *Teaching oral communication in grades K-8*. Boston: Allyn & Bacon.
- Creswell, J. M. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.
- Davar, N. F. , Dewan, M. A. A, & Zhang, X. (2025). AI chatbots in education: Challenges and opportunities. *Information 2025*, 16, 235. <https://doi.org/10.3390/info16030235>
- Duong, T. V. T. & Suppasetseree, S. (2024). The effects of an Artificial Intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning, Teaching and Educational Research*, 23(3), 293-321. <https://doi.org/10.26803/ijlter.23.3.15>
- Firman, F. & Haq, M. Z. (2012). Improving the students' speaking accuracy through "Ise 9.0 software version" (A classroom action research at the first year electric of SMK Negeri 1 Pattalassang Gowa). *Exposure Journal*, 2(1), 166-184. <https://doi.org/10.26618/ejpb.v1i2.777>
- Gokcearslan, S., Tosun, C., & Erdemir, Z.G. (2024). Benefits, challenges, and methods of Artificial Intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education (IJTE)*, 7(1), 19-39. <https://doi.org/10.46328/ijte.600>
- Golzar, J., Noor, S & Tajik, O. (2022). Convenience sampling. *International Journal of Education & Language Studies*, 1 (2), 72-77. <https://doi.org/10.22034/ijels.2022.162981>
- Grab, M. O. (2025). Integrated AI chatbot practice: A pathway to improved ESL speaking skills. *Social Sciences & Humanities Open*, 12, 101933. <https://doi.org/10.1016/j.ssaho.2025.101933>
- Halimah, A. N., Budianto, A., Hoesny, M. U., Cabuquin, J. C., & Ulfa, S. (2025). Improving English speaking skills through AI-based voice chatbots: A systematic review of vocational education studies. *Studies in Learning and Teaching (SiLeT)*, 6(3), 785-797. <https://doi.org/10.46627/silet.v7i1.684>
- Harmer (2001). *The practice of English Language Teaching*. London: Longman.
- Heriansyah, H. (2012). Speaking problems faced by the English department students of Syiah Kuala University. *Lingua Didaktika*, 6(1). <http://doi.org/10.24036/ld.v6i1.7398>
- Husnawati (2017). *Students' speaking performance: some challenging factors* (unpublished thesis). UIN Ar-Raniry, Darussalam- Banda Aceh.
- Koizumi, R. (2005). Speaking performance measures of fluency, accuracy, syntactic complexity, and lexical complexity. *JABAET (Japan-Britain Association for English Teaching) Journal*, 9, 5-33.

- Kurum, E. Y. (2016). Teaching speaking skills. In E. Solak Editor, *Teaching language skills for prospective English teachers* (pp.45-64). Pelikan.
- Le, N. Q. A., & Ngo, T. C. T. (2025). Enhancing presentation skills through Artificial Intelligence: A case study at Van Lang University. *International Journal of AI in Language Education*, 2(3), 1-18. <https://doi.org/10.54855/ijaile.25231>
- López-Molines, L. (2025). A systematic review of Generative Artificial Intelligence-based tools to improve oral skills in English as a Foreign Language. *The EuroCALL Review*, 32(2), 183-195. <https://doi.org/10.4995/eurocall.2025.23941>
- Martin, T. (2025). The emergence of AI chatbots in education. *Journal of Graduate Education Research*, 6(2), 49-58. Retrieved from <https://scholarworks.harding.edu/jger/vol6/iss2/7/>
- Meenakshi, R., & Sharma, B. (2026). Benefits and drawbacks of AI in education: A conceptual note. *Shodh Manjusha: An International Multidisciplinary Journal*, 3(1), 28–36. <https://doi.org/10.70388/sm250186>
- Tuan, N. H. & Mai, T. N. (2015). Factors affecting students' speaking performance at Le Thanh Hien High School. *Asian Journal of Educational Research*, 3(2), 8-23.
- Pitts, G., Marcus, V., & Motamedi, S. (2025). Student perspectives on the benefits and risks of AI in education. *2025 ASEE Annual Conference & Exposition*. <https://doi.org/10.48550/arXiv.2505.02198>
- Rachmadani, K. O., Musyafa'ah, I. R., Januanto, A. H., & Fitriyah, I. (2026). Exploring the impact of AI chatbots on students' speaking skills in EFL learners: A systematic review. *Eternal: English Teaching Journal*, 17(1), 229-240. <https://doi.org/10.26877/eternal.v17i1.3208>
- Safitri, E. I., Hidayati, S., & Ciptaningrum, D. S. (2025). The impact of AI chatbots on English language learners' speaking proficiency: A systematic review. *Journal of Research on English and Language Learning (J-REaLL)*, 6(2), 317–329. <https://doi.org/10.33474/j-reall.v6i2.23866>
- Thornbury, S. (2005). *How to teach speaking*. UK: Pearson Education Limited.
- Torky, S. A. E. F (2006). *The Effectiveness of a task- based instruction program in developing the English Language speaking skills of secondary stage students*. Ph.D. thesis (unpublished). Ain Shams University.
- Tran, T. T. T. (2025). The impact of AI-powered speaking tools on EFL students at Nguyen Tat Thanh University. *MSI Journal of Arts, Law and Justice*, 2(6), 1-20. <http://zenodo.org/records/15675420>
- Zhang, W. (2026). The impact of AI chatbots on EFL learners' oral proficiency and willingness to communicate. *System*, 136, 103919. <https://doi.org/10.1016/j.system.2025.103919>

Biodata

Ngoan Pham is currently an official lecturer at Nguyen Tat Thanh University, Vietnam. She has five years of experience in teaching students at the tertiary level. She has been in charge of

instructing both English-majored and non-English-majored students. Her research interests include methodology, CALL, AI application in English teaching, and testing and assessment.

Assoc. Prof. Dr. Pham Vu Phi Ho, Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam, was previously Vice President of Ba Ria – Vung Tau University and Vice President of Van Hien University, Vietnam. He created and designed [ICTE Robots](#) for language learners and IELTS practice. Pham has published more than 80 research articles in both local and International Journals, such as Computer-Assisted language Learning, SAGE Open, System, and Journal of Psycholinguist Research (25 Scopus/15 ISI), and 10 books and course-books, 3 edited books (2 Scopus-indexed) published by Springer Nature (2) and IGI Global (1) of which 3-course books were used for both the undergraduate and graduate levels at Van Lang University, HCMC Open University, Vietnam, and Lourdes College, Higher Education Department, Cagayan de Oro City, Philippines. He has international *teaching experience at Suranaree University of Technology in Thailand and Gyeongju University in South Korea*. He is the President of AsiaCALL. He is now the Editor-in-Chief of the International Journal of TESOL & Education. He is a managing editor of CALL-EJ (Scopus Q1), an editorial staff member for the [Suranaree Journal of Social Science](#) (Scopus Q3), and a peer reviewer for some international Journals indexed in ISI/Scopus, such as Computer Assisted Language Learning, Open Sage, Elsevier, and Frontiers ([Scopus](#)), ([ISI](#)), ([Google Scholar](#)).

Effectiveness of ChatGPT Feedback on Low-Level English Students' Writing Proficiency and Their Perceptions: An Experimental Study

Vu Quynh Hoa^{1*}, Nguyen Kim Phuc¹

¹ Faculty of Foreign Languages, Hanoi University of Science and Technology, Vietnam

*Corresponding author's email: hoa.vuquynh@hust.edu.vn

 <https://orcid.org/0009-0008-4785-1132>

 <https://doi.org/10.54855/ijaile.26324>

® Copyright (c) 2026 Vu Quynh Hoa, Nguyen Kim Phuc

Received: 22/06/2026

Revision: 16/08/2026

Accepted: 20/08/2026

Online: 31/08/2026

ABSTRACT

Keywords:

ChatGPT, Writing skills, Low-level students

The purpose of this study is to explore the effectiveness of ChatGPT's feedback on the writing of non-English major university students and their perceptions of this implementation. Participants in this study had a low level of English (A2). Researchers employed an experimental research design to achieve the research objectives. There were 42 students in total, including 21 students in the experimental group (EG) and 21 in the control group (CG). EG received ChatGPT feedback after writing draft 1, whereas CG received peer feedback and general instructions from teachers. Data were collected through pre- and post-tests, and semi-structured interviews. Paired sample t-tests revealed a significant improvement in mean scores between pre-and post-tests of EG ($p=.00<.05$), in contrast to CG's. While students complimented ChatGPT on its immediacy, dialogic, and personalized responses, the tool's prompting difficulties and blind acceptance were their major concerns. Findings suggest the need for clear AI literacy training for students and a 'human-in-the-loop' approach in which ChatGPT feedback incorporates teachers' guidance.

Introduction

In recent years, the use of ChatGPT to provide formative feedback in writing instruction has gained considerable attention from researchers. Studies show its potential in improving learners' writing proficiency (Abdi et al., 2024; Lai & Trinh, 2025; Zhang & Chen, 2026). Specifically, for learners, ChatGPT provides instant, detailed, and personalized feedback; meanwhile, for teachers, ChatGPT, like other AI generative tools, is regarded as a productive teaching assistant that helps teachers give learners timely feedback and reduces workload in large classrooms (Ardito, 2024).

Since its public release, there has been plenty of research studying the use of ChatGPT in education in general and teaching English writing in particular. When conducting research into ChatGPT's effectiveness in giving writing feedback, the researchers focused on recent, reliable, and insightful studies with similar objectives. In general, studies confirm ChatGPT's potential in giving students automated feedback.

Across these studies, ChatGPT has been associated with measurable gains in writing scores, grammar accuracy, thesis development, idea organization, and expression (Meniado et al., 2024; Poláková & Ivenz, 2024), particularly when combined with teacher feedback (Asadi et al., 2025) or applied to specific components of writing such as grammar, coherence, and vocabulary (Wang et al., 2023; Chan & Hu, 2023). At the same time, comparative and qualitative studies caution that ChatGPT still falls short of human feedback in analytical depth, personalization, and context-sensitivity (Steiss et al., 2024; Ljubojević, 2024). A more recent Vietnamese study further found that ChatGPT tends to assess student writing differently from teachers, often assigning notably higher scores across rubric criteria, with the greatest discrepancy occurring in content evaluation (Ho et al., 2026). This pattern is corroborated by a broader review of ChatGPT's role in written corrective feedback, which similarly found that while the tool offers benefits such as immediacy and clarity, its effectiveness for advanced writing skills, rhetorical elements, and cultural nuance remains limited compared to teacher feedback (Nguyen, 2026).

Despite promising results, a closer look at the studies reviewed above reveals a gap: none specifically targeted low-level English language learners, and their focus remained mainly on writing tasks rather than free writing. Consequently, how ChatGPT's feedback supports low-level learners in free writing tasks remains largely unexplored. The research team conducted this study to provide more empirical evidence for the role of ChatGPT in giving learners' writing formative feedback, simultaneously clarifying its role in supporting sustainable and autonomous learning skills.

Literature Review

Formative feedback in writing

Formative feedback in writing is generally understood as comments and suggestions on a draft, aimed at helping writers revise it to improve their products (Texas A&M University, 2026). Formative feedback plays a key role in improving learners' writing ability. According to Steiss et al. (2024), formative feedback in the writing process helps students understand what constitutes a quality piece of writing and how to improve different aspects of their writing. This, hence, helps plan suitable actions and effective learning strategies (Graham et al., 2011; Panadero & Jonsson, 2013). Therefore, formative feedback is not only a supporting tool for writing skills but also nurtures independent learning skills and critical thinking for learners (Lam, 2018; Mahapatra, 2024). So far, two types of formative feedback have been recognized. They are (1) traditional formative feedback, which is completely given by human beings (teachers/ peers), and (2) automated feedback, which is given by machines.

However, in practice, conventional formative feedback, which requires personalization for learners, takes considerable time and effort, leading many teachers to hesitate or refuse to do the task. This calls for solutions that reduce teachers' workload while still ensuring quality formative feedback for learners. Effective formative feedback, according to Steiss et al. (2024), includes clarity, precision, and a focus on the important points in the writing, with supportive tone. Hattie & Timperley (2007) claimed that students' motivation might increase, thereby improving writing skills, thanks to supportive, non-directive feedback in a supportive tone. These requirements were somewhat addressed by automated writing feedback, as students could receive fast, constructive feedback, which helped learners improve their writing results significantly (Fleckenstein et al., 2023).

Formative automated writing feedback

However, it was noted that early automated writing feedback tools had limitations, including low precision and the need for complex data training (Shermis & Wilson, 2024). In this context, the development of AI tools such as ChatGPT brings new opportunities because it could resolve the limitations above (Kurt and Kurt, 2024). For example, according to Steiss et al. (2024), ChatGPT could provide more flexible, precise, and personalized responses. Besides, Asadi et al. (2025) claimed that ChatGPT could also stimulate comprehensive development in language learning and students' motivation.

Research by Poláková and Ivenz (2024) shows that writing skills, distinct from other English skills, require formative feedback and repeated practice. Therefore, ChatGPT, with its instant and continuous responses, has proven its value in promoting learners' progress. Besides, Poláková & Ivenz (2024) claimed that formative feedback is commonly used in self- and peer-assessment, thereby promoting reflection, self-regulation, and self-correction in students' writing (Lam, 2018). With this idea, automated feedback has become an important part of the automation and upgrade of formative feedback. Studies by Rahman et al. (2022) show that students respond positively to automated feedback for identifying grammatical errors and improving writing skills. Automated feedback not only reduces teachers' burden (Wilson et al., 2022) but also contributes to learners' self-regulation (Grimes & Warschauer, 2010).

However, it is noted that ChatGPT should be used as a supporting tool, complementing teachers' feedback rather than completely replacing feedback sources (Chun et al., 2021; Zhang & Hyland, 2018). One notable reason was ChatGPT's misidentification of students' mistakes due to its preconfigured prediction algorithm (Hasanah & Safitri, 2024). The combination of AI-automated feedback and insightful feedback from teachers promises to create a flexible learning environment that develops writing skills in a sustainable and more rounded way.

ChatGPT prompting techniques

Designing an effective prompt for ChatGPT to provide formative writing feedback to low-level English students is key to exploiting this AI tool's potential. This group of students often struggles with presenting ideas, organizing essays, and applying correct grammar (Firkins et al., 2007; Aldabbus & Almansouri, 2022). Therefore, formative feedback needs to be clear, simple, and step-by-step. To achieve this, prompting techniques introduced by Brookman (2025, as cited in Stillman, 2025), which considered CLEAR (Lo, 2023), IDEA (Google for Educators, 2024), and the zero-shot prompting technique (Chae & Davidson, 2025), have been taken into account for the purpose of suggesting well-made prompts for students.

CLEAR (Lo, 2023) emphasizes 'concise, logical, explicit, adaptive and reflective' in prompts. Meanwhile, IDEA (Google for Educators, 2024) adds personalized factors for prompts through 'personal role, aim, recipients, theme, and expected structure'. This is particularly crucial in the context of learning to write, as feedback should be precise and suitable for learners' level. In addition, zero-shot, which means providing only questions without training the tool or sample examples, was chosen to introduce to students in this research to simplify the interaction process. Finally, the prompting technique outlined by Brookman (2025, as cited in Stillman, 2025) could be considered a practicalization of the theoretical frameworks discussed, including CLEAR and IDEA. Therefore, essential factors in the prompt include: (1) state your goal, (2) specify your preferred format, (3) warnings and guardrails, and (4) context dump. Beyond these structured frameworks, Woo et al. (2023) observed that non-technical users, including EFL students, often approach prompt writing through a trial-and-error process, gradually refining their prompts based on the chatbot's responses.

To sum up, formative feedback has long been recognized as a key factor in improving EFL learners' writing skills. However, formative feedback that meets essential criteria, including clarity, precision, and personalization, is time-consuming and effort-intensive for teachers. While the development of AI tools such as ChatGPT has shown potential in reducing teachers' workload by providing instant and personalized feedback, some scholars emphasize that ChatGPT should serve as a supporting tool for teachers in giving writing feedback rather than completely replacing them. At the same time, finding a suitable prompt in this context for low-level English students is essential for assessing the effectiveness of ChatGPT's feedback in enhancing students' writing skills. With this existing literature, we aim to explore (1) the effectiveness of ChatGPT feedback on low English level students' writing ability compared to traditional ways of giving feedback in our teaching context; and (2) students' attitudes towards the experience of working with ChatGPT in learning writing.

Research Questions

The study aims to address two research questions:

1. How did ChatGPT feedback affect low-level English students' writing ability compared to traditional ways of giving feedback?
2. What were students' attitudes towards ChatGPT feedback applied in their English writing lessons?

Methods

Pedagogical Setting & Participants

The research was conducted in two General English classes in a science and technology university that one of the two researchers taught during the second term of the 2024-2025 school year. In total, students studied 6 units, including reading, listening, communication skills, and writing lessons. Therefore, all 6 writing lessons of the 6 units in EG received intervention. Before the intervention, all participants were informed about the intervention by the teacher, who was also the researcher. Each of them then received an informed consent form. After receiving confirmation of participation, the researchers proceeded to the next stages of the research.

All students specialized in technical fields; they studied English as a mandatory subject in the first year of university. A convenience sampling method was chosen; students in both classes were at the A2 level in the CEFR, and class sizes were roughly the same. EG had 24 students, and CG had 23 students; however, two students in EG were excluded because of their absence in many lessons, and one student in CG was not suitable for the test because that student's English level was B2 (he was in class because he did not show up in the replacement test). Therefore, 21 students in each group participated in the study.

Design of the Study

Figure 1.

Research design

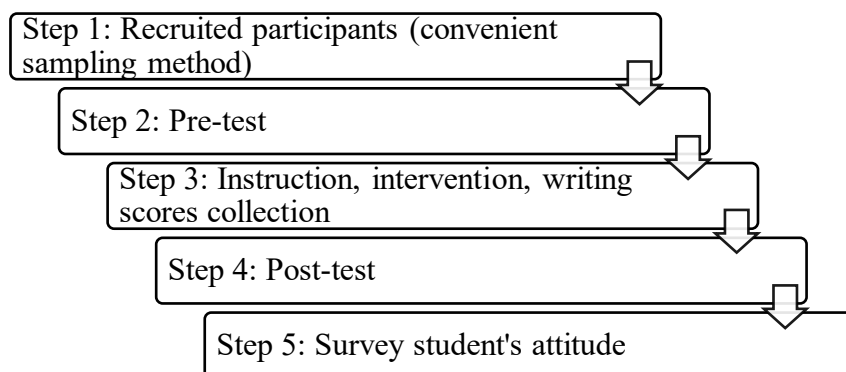


Figure 1 maps the research design. To achieve the research aims, the researchers first selected participants using convenience sampling, as they were the two most suitable and accessible groups. Then, in the second step, students were asked to take pretests to assess their initial writing proficiency. In the next step, the teacher gave EG students instructions on using ChatGPT feedback for their writing compositions, then began the intervention and collected writing scores. The details of the intervention process are presented in the next paragraph. In the fourth step, researchers administered post-tests to assess changes after the intervention, and, finally, we surveyed students' attitudes towards the intervention through semi-structured interviews.

The intervention was conducted as follows. In writing lessons, the teacher followed a step-by-step process: pre-writing, teaching students to write first drafts, providing feedback, and then revising drafts. For EG, after writing the first drafts and receiving the teacher's instruction, students used ChatGPT to get feedback; they then studied the feedback and wrote revised drafts based on ChatGPT's comments. Meanwhile, CG learned using traditional methods, which means they received peer feedback and learned from teachers' feedback for some random compositions in class. In total, there were six writing lessons.

After the intervention, five students from EG were selected to participate in the semi-structured interviews based on their availability and their writing scores in both the pre-test and post-test. We also aimed to ensure diversity among the participants by selecting individuals with a wide range of scores - from the lowest to the highest. Specifically:

Table 1.

Information of participants taking part in semi-structured interviews

Student code	Gender	Pre-test score	Post-test score
S1	Female	Lowest - 39	Lowest - 55
S2	Male	Low - 42	Low - 57
S3	Male	Medium - 58	Medium - 71
S4	Female	Medium - 58	Medium - 68
S5	Male	Highest - 60	Highest - 82

Data collection

In this research, we collected both quantitative data and qualitative data. Quantitative data were students' writing scores, which were gathered through pre-tests, post-tests, and three randomly

revised writing drafts in three lessons. Qualitative data, which were students' accounts of their attitudes towards the intervention experience, were gathered through semi-structured interviews. To collect data, an interview guide was designed based on open-ended questions covering two broad aspects: (1) students' attitudes towards the experience using ChatGPT to get formative feedback regarding its usefulness and ease of use, and (2) students' experience with generating prompts for ChatGPT feedback, including the language and procedures they used, the difficulties they encountered, and their suggestions for applying ChatGPT feedback effectively in writing lessons.

Data analysis

The quantitative data were students' writing scores; the researchers used Jacobs et al.'s (1981) ESL composition profile as an official, reliable scoring rubric to evaluate students' writing in both CG and EG. This was because the scoring rubric was claimed to provide a comprehensible evaluation of writing products, especially for teaching and scoring ESL learners. (Setyowati et al., 2020). Scores of pre-tests, post-tests, and three random pieces of writing were then compared by an independent t-Test. Regarding qualitative data, students' attitudes and responses from five students across three levels of performance in EG were transcribed and analyzed for themes.

Results and Findings

Research question 1: How did ChatGPT feedback affect low English level students' writing ability compared to traditional ways of giving feedback?

To evaluate the effectiveness of the intervention, the mean, SD, and p-value from an independent-samples t-test were used to compare the average scores of the Control Group (CG) and Experimental Group (EG) at different time points.

Table 2.

Results of independent t-tests (pre-test)

Pre-test	Mean	Std. Deviation	p-value
CG (n = 21)	50.76	9.48	> 0.05
EG (n = 21)	51.90	7.02	

Before the intervention, pre-tests showed little statistical difference between the two groups. The mean score of CG was 50.76 (SD = 9.48), while EG's was 51.90 (SD = 7.02). $p > 0.05$ indicates no statistically significant difference between the two groups. This confirms the similarity in writing ability between the two groups at the beginning of the study, paving the way for evaluating the effectiveness of the intervention.

Figure 2.*Results of three pieces of writing (revised drafts)*

The researchers selected three writing lessons to compare students' writing progress during the intervention. In writing 1, EG reached 70.9 for mean score (SD = 9.75), as opposed to CG's mean score of 55.61 (SD = 9.42), p -value < 0.05. This difference indicates statistical significance. In the second and third writings, the trend continued, with p < 0.05, affirming the positive impacts of ChatGPT's feedback on students' writing. In the third piece of writing, although EG's mean score was almost the same as the previous writing, the difference with CG was stable, with statistical significance. This indicates that EG tended to reach a 'threshold point', while CG continued to show a slow but steady increase.

Table 3.*Results of independent t-tests (post-test)*

Post-test	Mean	Std. Deviation	p-value
CG (n = 21)	60.81	9.08	< 0.05
EG (n = 21)	69.85	8.66	

Table 3 shows the post-test results for two groups, confirming an overall trend. The mean score of CG was 60.81 (SD = 9.08), and EG's mean score was 69.85 (SD = 8.66), p < 0.05. The difference is statistically significant, indicating that the intervention considerably improved students' writing ability compared to traditional teaching methods.

Table 4.*Results of independent t-tests regarding improvement in mean scores between pre-and post-test*

Group	Pre-Test	Post-Test	p-value
CG (n = 21)	50.76 ± 9.48	60.80 ± 9.08	< 0.05
EG (n = 21)	51.90 ± 7.02	69.85 ± 8.66	< 0.05

Table 4 reveals a significant improvement in mean scores between pre-and post-tests of EG ($p = .00 < .05$), in contrast to CG's.

Research question 2: What were students' attitudes towards ChatGPT feedback applied in their English writing lessons?

All participants generally shared positive thoughts towards ChatGPT feedback, claiming that it was prompt and detailed. As S5 explained,

Unlike humans, the software doesn't care whether its responses hurt or offend users. It simply points out the problems directly and specifically. That's especially helpful for students with lower English proficiency because they have many knowledge gaps and need clear feedback to improve their writing.

All five interviewees also affirmed that ChatGPT feedback helped them fix grammar, vocabulary, and spelling errors. In particular, this tool helped improve students' "use of verb tenses" (S4), provide grammatical knowledge that students had misunderstood or had not yet known (S2), and "suggest more accurate word choices based on the given contexts" (S3). Furthermore, ChatGPT feedback was said to enhance students' content organization. According to S2 and S4, by helping them restructure their writing outlines, this AI tool not only turned the sketchy original drafts into more logical ones but also improved their thought flow. For S5, the most significant strength of ChatGPT is its ability to present information logically and coherently. This allows students with low English levels to imitate it to strengthen their reasoning.

However, the tool still contains certain weaknesses. As stated by S3 and S4, although ChatGPT could often provide specific and prompt evaluation of writing quality, students themselves had to consider whether the issues raised by this software were valid. For instance, sometimes ChatGPT "falsely identified the verb tense error" as it misunderstood the context (S4), or "sometimes it would provide too general feedback that I could not identify what the error was" (S2). This could be challenging for students with low English proficiency, since they tend to make many errors and may not recognize all of them. Besides, this AI tool could fail to answer students' questions correctly and even provide confusing information in its responses (S5). Notably, when S1 asked ChatGPT to evaluate her use of punctuation and capitalization, the software completely ignored the request, focusing solely on analyzing her grammar and vocabulary errors. This may hinder students from developing comprehensive writing skills, as the software overlooks important criteria they must follow.

Students' prompt-generating experience

a. Language use for generating prompts

Most participants utilized Vietnamese during their conversations with ChatGPT. Explaining this, S3 shared that writing prompts in the mother tongue helped him express requests more clearly, and ChatGPT could also rely on that to respond more appropriately. For S1, her current low level of English made it impossible for her to express her thoughts in that language. Therefore, she chose to use Vietnamese to communicate with the tool instead. She also argued that

if I thoroughly inquire and know how to harness AI data effectively, then even if I interact with ChatGPT in Vietnamese, I can still collect useful knowledge and thereby enhancing my English writing proficiency.

Among all participants, only S5 used a mixture of Vietnamese and English to communicate with ChatGPT. However, he stated that when he mixed both languages in particular prompts, ChatGPT tended to respond in English only. His solution was to keep reading the AI feedback.

If he did not understand anything, he would ask the software to simplify or translate such information into Vietnamese.

b. Prompt generating procedure

Although the general prompt-generating procedure had been taught in class, participants affirmed that they gradually adjusted and tailored it to make it more convenient for them.

Differences in the prompt-generating procedure among participants can be illustrated in the following step-by-step workflows:

Figure 3.

The prompt-generating procedure of S3 and S4

Step 1: Provide basic requirements

- Provide feedback for my writing. The word limit required is ... words
- The writing question is ...
- Provide feedback on grammar, vocabulary and content of the writing
- What do I need to improve?
- I am at A2 level, so please respond using A2-level English.

Step 2: Provide the full writing

Step 3: Receive AI feedback. Use follow-up prompts for deeper insights

- Clarify this error. Why is it an error and how do I fix it?
- You noted this part as problematic. How is it so?
- How to further develop the content of this part?

Change this word into a more appropriate one

Figure 4

The prompt-generating procedure of S1 and S2

Step 1: Provide the full writing immediately

Step 2: Notify ChatGPT that I am at A2 level, so it should respond accordingly

Step 3: Request an overall score

Step 4: Ask ChatGPT to provide more specific feedback related to grammar, vocabulary and content of the writing

Figure 5

The prompt-generating procedure of S5

Step 1: Self-evaluation- Review the writing carefully, take notes on identified weaknesses

Step 2: Provide basic requirements, also ask ChatGPT to focus on weaknesses that were identified

Step 3: Provide the full writing

Step 4: Receive AI feedback. Use follow-up prompts for deeper insights, especially focusing on targeted areas to improve writing skill

S5 stated,

Writing ability comes from both talent and practice. After practicing writing English for a while, I believe I should only focus on my weaknesses and use ChatGPT to improve them in the most effective way. This not only saves time but also enhances self-learning efficiency. If I simply share the whole writing and ask ChatGPT to correct it, the results I receive would be too vague- the software provides too much information, making it difficult to filter out the most relevant and necessary points to me.

c. Challenges in prompt generation

S1 complained that ChatGPT often automatically rewrote her entire piece of writing, even though she only needed it to provide feedback and fix her errors. Besides, ChatGPT sometimes did not understand prompts provided by students. Consequently, its responses did not meet their expectations (S2, S3 and S4).

Thanks to the considerable time spent interacting with ChatGPT, all participants gained experience dealing with such issues. For instance, S2 would simplify his prompts using simpler lexis and shortened expressions. S3 gave ChatGPT an example to illustrate his point clearly before asking the tool to respond.

With S4, her problem was that ChatGPT sometimes forgot what her English level was and thereby provided feedback with much more complex language: *“Sometimes, ChatGPT provides very advanced sentences with lots of words and structures that I do not understand”* (S4). To solve such a problem, she had to keep reminding the software of her English proficiency.

Students’ suggestions to apply ChatGPT feedback for effective writing

S2 believed ChatGPT should be used during the Pre-writing stage to aid in brainstorming and structuring content, especially since he struggled to organize ideas due to limited writing experience. He suggested that if ChatGPT generated initial content in Vietnamese, it could help him apply grammar and vocabulary more effectively. In contrast, S4 and S5 preferred to use ChatGPT during the Drafting and Revision-Feedback stages, after students had already developed their writing independently. S5 strongly opposed its use in Pre-writing, arguing that early reliance on AI could hinder students’ original thinking.

All participants planned to continue using ChatGPT feedback for their writing practice but emphasized the need to receive evaluations from both the software and their teachers. S2 shared that while teachers provide tailored feedback based on each student’s proficiency and learning needs, ChatGPT could help support large classes, especially when there is not enough in-class time for feedback. S5 agreed, noting that while ChatGPT’s polished style can feel mechanical and difficult to understand, teachers suggest using accessible, natural language for weaker students. For S4, since this AI tool sometimes provides too general feedback, students must actively seek additional information to gain insights into certain aspects of their writing, for example, by crafting specific prompts to receive responses tailored to their needs. In contrast, teachers provide comprehensive feedback on all aspects of students’ writing, allowing them to simply listen, reflect, and improve without spending extra time searching for information. This makes the learning process more comfortable and efficient. To save time and ensure that teachers can still review each student’s work, S2 and S3 suggested that students should first share their writing with ChatGPT for initial feedback. Afterward, they can share such AI-generated comments with their teachers, allowing the teachers to evaluate their accuracy and provide additional feedback to further refine the writing.

Discussion

Quantitative data indicate that students who received ChatGPT's feedback showed an apparent improvement in writing ability compared to those who didn't. This result is similar to previous studies about automated writing feedback tools in teaching English writing. For instance, Mahapatra (2024) reported that ChatGPT, as a generative feedback tool, had positive effects on ESL students' academic writing ability. Systematic reviews also reported that automated writing feedback can provide constructive feedback, leading to significant improvements in students' writing (Fleckenstein et al., 2023). These findings affirm that the use of AI-based technology such as ChatGPT can help students enhance their writing skills by providing timely and personalized responses. Research findings are also similar to surveys on ChatGPT's effectiveness in language education. For example, Kurt and Kurt (2024) found that students accepted ChatGPT for its ease of use and useful responses. Simultaneously, Kurt and Kurt's (2024) study warned that ChatGPT can only solve linguistic problems on the surface and sometimes provides inconsistent answers; therefore, it is impossible to completely replace teachers' or peers' feedback with ChatGPT's. Whether participants in this research provided similar opinions will be presented in the discussion for the second research question.

Regarding the effectiveness of ChatGPT feedback, the qualitative findings indicate that this tool enhanced students' writing skills, particularly in grammar, vocabulary, and content organization. This is similar to Ljubojević (2024), in which ChatGPT was found to help students improve their understanding of the structure and genre of their writing, generate ideas, and enhance their English language skills.

However, providing inaccurate feedback is one of the tool's typical weaknesses. For our study, participants realized that ChatGPT sometimes incorrectly detected grammatical errors. Similarly, as stated in Hasanah & Safitri (2024), ChatGPT's misidentification of the student's mistake suggests that, rather than understanding the given text, this software relies on a predictive algorithm to generate its response. Notably, this automated writing evaluation tool could be overly general and wordy and may at times confuse those receiving feedback (Grimes & Warschauer, 2010; Shermis & Wilson, 2024). This is also consistent with our interviewees' opinions. According to Grimes & Warschauer (2010), inaccurate or confusing feedback can discourage students, especially novice writers, from engaging in revision. To address this problem, Hasanah & Safitri (2024) suggested that teachers remind students that AI-generated feedback is not always accurate, and that students should carefully consider its accuracy and usefulness before integrating it into their writing. Since students in our study had low English proficiency, teachers could scaffold their reflections.

Based on our findings, 2 students with the lowest pre- and post-test scores (S1, S2) used a different prompt-generating procedure than the 3 other students with higher pre- and post-test scores (S3, S4, S5). In particular, S1 and S2 initially provided ChatGPT with their writing and then generated additional prompts to elicit ChatGPT's feedback. This procedure reflects *trial-and-error prompting*, in which students adjusted their prompts in real time based on the responses received. They learned through repeated attempts which types of prompts best elicited the feedback they needed (Woo et al., 2023). Meanwhile, S3, S4, and S5 chose to provide ChatGPT with the essential requirements first, then shared their writing with the tool. The strategy applied here was *chain prompting* - a structured prompting strategy where users break down their requests into sequential steps, ensuring that AI responses are progressively refined and contextually relevant (Wei et al., 2022). Although there have been no studies directly proving the relationship between prompt-generating strategies and students' English proficiency, the differences observed in prompt-generating procedures suggest that higher-

proficiency students tend to engage in more structured approaches like chain prompting, as they have a clearer understanding of their needs and expectations, whereas lower-proficiency students often engage in trial-and-error prompting – adjusting their requests based on AI responses. This implies that, in addition to language ability, to apply AI feedback effectively in English writing, students need to develop strategic thinking and cognitive flexibility when crafting prompts. Teachers may consider providing students with further guidance and practice on these elements, along with structured prompting techniques, so they can better control AI response quality, especially since prompting affects the type of feedback generated (Steiss et al., 2024).

Although S5 appreciates ChatGPT's straightforward approach to feedback, which offers thorough comments on low-level students' writing issues, Hattie & Timperley (2007) suggested that students may feel more motivated and improve their writing skills more effectively when receiving feedback in a *supportive, non-directive* tone – similarly to what is typically provided by human teachers. Although this has not been verified in the study, apart from S5, no other participants reported that the straightforward versus supportive tone of feedback affected their writing skills; students generally expressed a strong preference for receiving feedback from teachers alongside ChatGPT. Some even noted that teachers provide deeper, more tailored insights, making their feedback more suitable for individual students than ChatGPT's. This is similar to Steiss et al. (2024), who indicated that human evaluators, particularly those who are well-trained, compensated, and have sufficient time, deliver higher-quality feedback in terms of clarity in guiding improvements, accuracy, and prioritization of essential aspects.

In our study, S5, a higher-proficiency student, opposed applying ChatGPT at the Prewriting stage, emphasizing independent idea generation and critical thinking. This is consistent with Hasanah & Safitri (2024), who stated that dependence on technology from the outset could diminish creativity and critical thinking. As students increasingly use AI-powered tools for idea generation and writing, they may rely less on their own cognitive processes. Over time, this reliance could weaken their ability to think independently and innovate. However, S2 – a lower-proficiency student – was supported by ChatGPT during the Prewriting stage because of his limited writing experience and strong desire to learn from AI-generated sample content. These contrasting experiences suggest that writing proficiency significantly shapes both the approach and extent to which students utilize ChatGPT feedback. It also influences whether they use AI as a refinement tool or rely on it for foundational guidance.

Participants' suggestion to combine feedback from both teachers and ChatGPT also aligns with Steiss et al. (2024), who recommend integrating AI-generated feedback for first drafts while reserving human evaluation for later or final revisions. This approach ensures that the human perspective remains central and allows instructors to address any shortcomings in AI feedback, such as inaccuracies or failure to emphasize critical areas for improvement.

Conclusion

This research has provided empirical evidence of the effectiveness of ChatGPT in providing formative writing feedback to low-level English learners. Quantitative data indicate that students who received ChatGPT's feedback significantly improved their writing skills compared to those in the CG. Simultaneously, qualitative data indicated that this tool could help learners refine their grammar, vocabulary, and organization. Additionally, they complimented this tool's immediacy, dialogic nature, and personalized responses. Regarding variations in prompting techniques (e.g., trial-and-error prompting and chain prompting), the study revealed

that higher-level students tended to approach ChatGPT more systematically and with a clearer goal, whereas low-level students were inclined to adjust their prompts based on the AI tool's responses.

However, the study also spotted some notable drawbacks of this means of feedback. Firstly, feedback was sometimes inaccurate or overly general, causing confusion among learners. Furthermore, the reliance on ChatGPT in brainstorming could reduce learners' critical thinking and creativity. Even though some learners appreciated ChatGPT's generated feedback, they still expected to combine teacher and ChatGPT feedback to ensure personalization and depth of comment. This highlights the necessity of teachers' role in supervising, supporting, and instructing the use of the AI tool.

This study has certain limitations. First, the sample size was small, and the study was limited to low-proficiency learners; therefore, the findings may not generalize to other types of learners. Secondly, there was no control over learners' interactions with ChatGPT while receiving feedback, making the effectiveness dependent on several factors, such as prompting ability, understanding of ChatGPT's responses, and learners' motivation.

The study suggests that future research should expand its scope and compare proficiency levels. Additionally, there should be more insightful studies on prompting strategies and the correlation between learners' understanding of AI feedback and their ability to optimize responses; all of these aim to enhance the implementation of ChatGPT and other generative AI tools in writing instruction.

References

- Abdi, M. M., Ananta, B. D. B., & Nurbatra, L. H. (2024). Exploring the use ChatGPT in improving students' paragraph composition skills. *Focus on ELT Journal*, 6 (4), 17-36. <https://doi.org/10.14744/felt.6.4.2>
- Aldabbus, S., & Almansouri, E. (2022). Academic writing difficulties encountered by University EFL Learners. *British Journal of English Linguistics*, 10(3), 1–11. <https://doi.org/10.37745/bjel.2013/vol10n3111>
- Ardito, C. G. (2024). Generative AI detection in higher education assessments. *New Dir Teach Learn*, n/a(n/a). <https://doi.org/10.1002/tl.20624>
- Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, 56, 101766. <https://doi.org/10.1016/j.tsc.2025.101766>
- Chan, C., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43), 1–18. <https://doi.org/10.1186/s41239-023-00411-8>.
- Chae, Y., & Davidson, T. (2025). Large Language Models for Text Classification: From Zero-Shot Learning to Instruction-Tuning. *Sociological Methods & Research*, 0(0). <https://doi.org/10.1177/00491241251325243>
- Chun, H. L., Lee, S. M., & Park, I. E. (2021). A systematic review of AI technology use in English education. *Multimedia-Assisted Language Learning*, 24(1), 87–103. <https://doi.org/10.15702/mall.2021.24.1.87>
- Firkins, A., Forey, G., & Sengupta, S. (2007). Teaching writing to low proficiency EFL students.

- ELT Journal*, 61(4), 341–352. <https://doi.org/10.1093/elt/ccm052>
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1162454>
- Google for Educators. (2024b). *Generative*. Mountain View, CA: AI for educators. Google LLC. <https://skillshop.exceedlms.com/student/path/1176018>.
- Graham, S., Harris, K., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment: A report from Carnegie Corporation of New York*. New York: Carnegie Corporation.
- Grimes, D. & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, 8(6). <https://www.ejournals.bc.edu/index.php/jtla/article/view/1625/1469>
- Hasanah, U., & Safitri, W. (2024). Student perception toward ChatGPT in writing skill in English study program Universitas Muhammadiyah of Bengkulu. *Teaching and Learning English in the Language Education (Telle)*, 4(2). <https://doi.org/10.36085/telle.v4i2.6937>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Ho, N. N., Tra, T. C. T., Pham, T. Q. & Tran, P. T. (2026). Discrepancies between ChatGPT and Vietnamese EFL teachers in writing assessment. *International Journal of AI in Language Education*, 3(1), 1–16. <https://doi.org/10.54855/ijaile.26311>
- Jacobs, H. L., Zinkgraf, S. A., Wormouth, D.R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House. [ERIC ED217708](https://eric.ed.gov/?id=ED217708). <https://eric.ed.gov/?id=ED217708>
- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, Article 24. <https://doi.org/10.28945/5370>
- Lại, T. K. L. & Trịnh, H. L. (2025). TÁC ĐỘNG CỦA CHATGPT ĐỐI VỚI NĂNG LỰC VIẾT HỌC THUẬT CỦA SINH VIÊN CHUYÊN NGÀNH TIẾNG ANH TẠI MỘT TRƯỜNG ĐẠI HỌC. *Tạp chí Nghiên cứu nước ngoài*, 41(3), 92-109. <https://doi.org/10.63023/2525-2445/jfs.ulis.5514>
- Lam, R. (2018). Feedback in writing portfolio assessment. In R. Lam (Ed.), *Portfolio assessment for the teaching and learning of writing*, (pp. 59–72). Springer. https://doi.org/10.1007/978-981-13-1174-1_5
- Ljubojević, D. (2024). Student feedback on the effectiveness of ChatGPT for essay writing. In *Proceedings of the 15th International Conference on e-Learning 2024* (pp. 116-126). CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3938/Paper_12.pdf
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(9). <https://doi.org/10.1186/s40561-024-00295-9>
- Meniado, J. C., Duong, T. T. H., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using

- ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence*, 7, 100313. <https://doi.org/10.1016/j.caeai.2024.100313>
- Nguyen, D. M. Q. (2026). The role of ChatGPT in providing written corrective feedback for EFL academic writing: A literature review. *International Journal of AI in Language Education*, 3(1), 35–47. <https://doi.org/10.54855/ijaile.26313>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11, 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- Rahman, N., Zulkornain, H., & Hamzah, H. (2022). Exploring artificial intelligence using automated writing evaluation for writing skills. *Environment-Behaviour Proceedings Journal*, 7(SI9), 547–553. <https://doi.org/10.21834/ebpj.v7iSI9.4304>
- Setyowati, L., Sukmawan, S., & El-Sulukiyyah, A. A. (2020). Exploring the use of ESL composition profile for college writing in the Indonesian context. *International Journal of Language Education*, 4(2), 171–182. <https://doi.org/10.26858/ijole.v4i2.13662>
- Stillman, J. (2025, February 25). How to write the perfect AI prompt, according to OpenAI President Greg Brockman. *Inc.com*. <https://www.inc.com/jessica-stillman/how-to-write-perfect-ai-prompt-openai-greg-brockman.html>
- Shermis, M. & Wilson, J. (Eds). (2024). *Routledge international handbook of automated essay evaluation*. Routledge (in press).
- 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, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Texas A&M University Writing Center. (2026). *Formative feedback*. Texas A&M University. <https://writingcenter.tamu.edu/faculty/resources/formative-feedback.html>
- Wang, M., & Guo, W. (2023). The potential impact of ChatGPT on education: Using history as a rearview mirror. *ECNU Review Education*. <https://doi.org/10.1177/20965311231189826>.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837. https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf
- Wilson, J., Myers, M. C., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, 1–17. <https://doi.org/10.1080/0969594X.2022.2025762>
- Woo, D. J., Guo, K., & Susanto, H. (2023). *Cases of EFL secondary students' prompt engineering pathways to complete a writing task with ChatGPT*. arXiv.

<https://doi.org/10.48550/arXiv.2307.05493>

Zhang, R. & Chen, J. (2026). The effects of ChatGPT on EFL students' writing skills: A meta-analysis. *System*, 140, 104043. <https://doi.org/10.1016/j.system.2026.104043>

Zhang, Z. (V.), & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>

Biodata

Ms. Nguyen Kim Phuc has been working as an English lecturer at the Faculty of Foreign Languages, Hanoi University of Science and Technology (HUST) since 2019. Her research in ELT primarily focuses on enhancing students' productive skills.

Ms. Vu Quynh Hoa has also been working as an English lecturer at the Faculty of Foreign Languages, Hanoi University of Science and Technology (HUST) since 2019. Her research interests are students' engagement, the application of technology, and EFL learners' writing skills.

International Journal of AI in Language Education

Vol. 3, No. 2 (2026): AI in language Education

Vol. 3, No. 2 (2026): AI in Language Education

Orcid: <https://orcid.org/0009-0004-5901-9176>

DOI: <https://doi.org/10.54855/ijaile.2632>

Publisher: ICTE Press

Address: International Association of TESOL & Education, 5243 Birch Falls Ln, Sugar Land, Texas, USA, 77479

About the Journal

Aims and scopes

The mission of the International Journal of AI in Language Education (IJAILE) (ISSN: 3065-0216) is to contribute to the literature of instruction by publishing high-quality studies in the areas of Artificial Intelligence, Computer-Assisted Language Learning (CALL), Computer Science Applications, Applications, AI Tools, Chatbots, m-learning, e-learning, Pedagogies & Language Teaching, AI for translation and interpretation, educational technology, and other Artificial Intelligence issues. As a peer-reviewed journal, one of our priorities is to ensure high standards of publishing. Publication ethics is an important concern, and editors, peer-reviewers, and authors should pay great attention to meeting ethical standards.

Open Access Policy

IJAILE aims to increase visibility and make easier use of open access scientific papers. Readers and their institutions are supported for online access without asking for any royalty, personal information, or log-in details. According to the open-access policy of our journal, all readers are permitted to read, download, copy, distribute, print, link to, and search our articles with no charge.

Copyright

The copyright of all articles published in the IJAILE remains with the Authors, i.e. Authors retain full ownership of their article. Permitted third-party reuse of the open access articles is defined by the applicable Creative Commons (CC) end-user license which is accepted by the Authors upon submission of their paper. All articles in IJAILE are published under the CC BY-NC 4.0 license, meaning that end users can freely share an article (i.e. copy and redistribute the material in any medium or format) and adapt it (i.e. remix, transform and build upon the material) on the condition that proper attribution is given (i.e. appropriate credit, a link to the applicable license and an indication if any changes were made; all in such a way that does not suggest that the licensor endorses the user or the use) and the material is only used for non-commercial purposes.

Ethics and Malpractice

The Statement of Publication Ethics & Malpractice

The International Journal of AI in Language Education (ijaile) (ISSN: 3065-0216) is committed to upholding ethical standards, retracting and correcting errors. The editorial team's primary responsibility is to discourage publishing malpractice. Any type of unethical conduct is unacceptable, and this Journal's Editorial Team does not tolerate plagiarism in any form. All manuscripts must be the authors' original work and free of indications of plagiarism.

Conflicts of Interests

Authors are requested to disclose interests that are directly or indirectly related to the work submitted for publication. Interests within the last 3 years of beginning the work (conducting the research and preparing the work for submission) should be reported. Interests outside the 3-year time frame must be disclosed if they could reasonably be perceived as influencing the submitted work. Disclosure of interests provides a complete and transparent process and helps readers form their own judgments of potential bias. This is not meant to imply that a financial relationship with an organization that sponsored the research or compensation received for consultancy work is inappropriate. Please download this form.

Disclosures and declarations

All authors are requested to include information regarding sources of funding, financial or non-financial interests, study-specific approval by the appropriate ethics committee for research involving humans and/or animals, informed consent if the research involved human participants, publication on vulnerable populations, and a statement on the welfare of animals if the research involved animals (as appropriate).

The decision on whether such information should be included is not only dependent on the scope of the journal, but also the scope of the article. Work submitted for publication may have implications for public health or general welfare and in those cases, it is the responsibility of all authors to include the appropriate disclosures and declarations.

Compliance with Ethical Standards

To ensure objectivity and transparency in research and to ensure that accepted principles of ethical and professional conduct have been followed, authors should include information regarding sources of funding, potential conflicts of interest (financial or non-financial), informed consent if the research involved human participants, publication on vulnerable populations, and a statement on the welfare of animals if the research involving animals.

Authors should include the following statements (if applicable) in a separate section entitled “Compliance with Ethical Standards” when submitting a paper.

Originality and Plagiarism

Authors can guarantee that their writings are entirely unique and that any writing or words credited to another are properly referenced. Additionally, publications that had an impact on the nature of the finding reported in the manuscript should be acknowledged. The writers must certify that the manuscript has never been written before.

It is not permitted to copy text from other publications without correctly attributing the source (plagiarism) or to publish several papers with nearly identical material by the same writers (self-plagiarism). It is against the law to concurrently apply the same results to more than one Journal. It is prohibited for authors to present results collected from others as their own. Authors should consider the work of those who shaped the direction and course of their research.

Check plagiarism

The IJAILE will conduct a plagiarism check on all submitted papers using Turnitin’s program. The ijte only considers papers with less than 20% similarity index for publication. IJAILE will immediately reject any paper that involves plagiarism.

AI and AI-Assisted Tools

Authorship implies human-only responsibilities and duties. Each author is responsible for ensuring that any questions regarding the accuracy or integrity of any portion of the work are adequately investigated and resolved, and authorship requires the ability to endorse the final version of the work and consent to its submission. The authors are also responsible for ensuring that the work is original, that the indicated authors qualify for authorship, and that the work does not violate the rights of third parties.

Papers generated by AI or AI-assisted Tools are not accepted for publication in this Journal. This policy pertains only to the writing process and not to the use of AI tools to analyze data and gain insights for the research procedure. This policy does not prohibit the use of artificial intelligence and AI-assisted instruments in formal research design or research methods. When AI or AI-assisted tools are used in this context, they should be described in the Methods section as part of the work's methodology.

The authors are responsible and answerable for the content of their work. The authors should indicate in their manuscript if they utilized AI and AI-assisted technologies, and a corresponding statement will appear in the final product. Disclosure of the use of these technologies promotes honesty and trust among authors, readers, reviewers, editors, and contributors and facilitates adherence to the tool's or technology's terms of service.

Duplicate submission

The authors must certify that the manuscript is not being considered for publication elsewhere during submission.

Corrections and retractions

All writers are required to notify and collaborate with journal editors to retract or correct errors in published works expeditiously.

The Journal will withdraw articles if there is compelling proof that the conclusions are false, either as a consequence of deception (data fabrication) or dishonest (miscalculation); or in case the results were published elsewhere without properly referenced, consent, or explanation (redundant publication); or the study involves plagiarism or immoral analysis.

Employers bear the primary responsibility for combating study fraud. We will check with referees and the Editorial Team if we become aware of any possible misconduct. If evidence exists, we will remedy the situation by making fair corrections to the online Journal; refusing to consider future work by an author for a given period of time, and contacting affected writers and editors of other journals.

Confidentiality

Authors should treat all communication with the Journal as confidential which includes correspondence with direct representatives from the Journal such as Editors-in-Chief and/or Managing Editors and reviewers' reports unless explicit consent has been received to share information.

Publication Ethics

The International Journal of AI in Language Education (ijaile) publishes papers that have been written within the framework of an appropriate research design. The relevant procedures must be followed. The editorial team and reviewers are tasked with the responsibility of monitoring these steps.

The editors and publishers are mindful of the ramifications of detecting research misconduct. In the event of ethical violation or research misconduct, ijaile removes the article from its website and requests an explanation from the corresponding author within seven days. The ijaile editorial team reviews the responses and makes the final decision on the report.

The Editorial Team and Publisher of the ijaile state unequivocally that any revisions, clarifications, retractions, or apologies will be published on the Journal's website.

Reviewers' responsibilities

Each reviewer makes a voluntary contribution to their time. Reviewers play a critical part in ensuring the editorial decision-making process is as high-quality as possible.

Reviewers must have an impartial assessment of the paper based on their existing knowledge of the field, as well as previous research and literature. They are required to complete their evaluations within the deadline specified when the request for review was accepted. This timeliness is critical in assisting writers in completing their manuscripts in a fair amount of time and in ensuring that journal issues are published on time. Reviewers are required to notify the publisher of any significant resemblance to other articles they are aware of. They are obligated to maintain the confidentiality of all analysis materials. Additionally, reviewers are required to notify the editor whether they feel they may be in a position to profit from the review (for example, but not limited to, declaring any professional or personal relationship which might compromise the review).

Decisions of publication

The editors ensure that all manuscripts submitted for consideration of publication receive peer review by at least two experts. The Editor-in-Chief or managing editor is responsible for selecting manuscripts for publication from those sent to the journal, taking into account the work's validity, value to academics and audiences, the views of critics, and all other applicable legal provisions including libel, copyright infringement, and plagiarism. When making this determination, the Editor-in-Chief or managing editor can consult with other editors or reviewers.

Brief Information:

Country of Publication: Texas, United States

Publisher: ICTE Press

Youtube: <https://www.youtube.com/@ictepress>

Open Access: Yes

ISSN: 3065-0216

Frequency: 3 issues per year

Publication Dates: April, August, and December

Language: English

Scope: Artificial Intelligence, AI Tools, Chatbots, Computer-Assisted Language Learning, Computer Science Applications, Applications, Educational Technology, and AI Issues in language education

Article Processing Charge: \$70

Types of Journal: Academic/Scholarly Journals

Indexed: Google Scholar, ORCID, Crossref, DOI, Library of Congress, Mendeley, PKP PN, LOCKSS, CLOCKSS, Semantic Scholar.

Policy: Double-blind peer review

Plagiarism check: Turnitin

Contact e-mail: editor@ijaile.org or editor.ijaile@gmail.com