



International Journal of Language Studies

ISSN: 3078-2244 (Online)

2026, Vol. 3, No. 2, pp. 105–125

DOI: <https://doi.org/10.60087/ijls.v3.n2.007>

Journal homepage: <https://ijlangstudies.org>



Beyond Conversation Practice: Investigating the Effects of Generative AI-Powered Voice Tools on EFL Learners' Speaking Fluency, Accuracy, and Willingness to Communicate

Zewudu Paulos Mada

Department of English Language and Literature, Jinka University, Ethiopia

*Corresponding author: Zewudu Paulos Mada, Email: zewudupaul4@gmail.com

Abstract

This study investigated the effects of the use of generative AI-driven voice tools on the fluency, accuracy, and willingness to communicate of EFL learners. Following a post-positivist paradigm and using a quasi-experimental pretest-posttest control group design, 30 EFL learners in their second or third year at Jinka University were included in a 12-week intervention program. While the experimental group had AI-driven speaking practice using ChatGPT with voice capabilities, the control group engaged in conventional peer-conversation activities. Statistical analyses showed that there were significant gains in the areas of speaking fluency, accuracy, and WTC for the experimental group compared to those for the control group. Moreover, qualitative findings suggest that speaking anxiety was decreased, increased confidence in communication was observed, and learners appreciated the low-stress practice experience.

Keywords: generative AI; speaking fluency; accuracy; willingness to communicate; voice tools; ChatGPT

Article information: Received: 21/03/2026;

Accepted: 25/05/2026;

Published: 25/06/2026

Copyright © 2026 The Author(s). Published by the International Journal of Language Studies. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

1. Introduction

1.1 Background of the Study

Incorporating artificial intelligence in language teaching is one of the revolutionary trends in the area of Teaching English to Speakers of Other Languages (TESOL). With regard to all the applications of artificial intelligence technology, the use of generative AI-enabled voice tools represents a revolutionary trend in the field of solving persistent problems in EFL speaking skills training. The implementation of these innovative solutions makes use of such sophisticated technological resources as natural language processing and speech synthesis that facilitate participation of students in real-life interactive conversations with AI-enabled conversational agents along with immediate feedback. In particular, according to Teaching English with Technology, the study of ways in which the combination of "human and AI" may enhance teaching and learning outcomes becomes a crucial issue of modern technology-enhanced language teaching research (Alexander & Krajka, 2025). At the same time, the journal encourages submission of manuscripts dedicated to investigation of the role of AI in TESOL in

classroom context and cultural/ethical issues of human-machine collaboration in English language teaching (Alexander & Krajka, 2025).

Speaking is one of the most difficult skills to master for EFL learners, especially in situations when there are no possibilities for interactions in a target language. Indeed, in traditional EFL classes, there is a high focus on reading and writing because of evaluation requirements; thus, there is very little practice of speaking. When speaking is practiced, there is a range of issues like the size of the group, students' fear of speaking, lack of individual practice, and difficulties with providing meaningful feedback to every student. All these issues pushed researchers and teachers to look for solutions in technology-based teaching methods which would complement classroom work and help to provide learners with more practice in speaking.

The use of voice generators powered by generative AI allows overcoming some of the issues associated with conventional methods of speaking practice. Generative AI-based tools allow providing learners with infinite practice of speaking in low-pressure settings and thus help to reduce the fear of speaking in classroom settings. Moreover, the tools allow providing instant feedback about different aspects of speaking like pronunciation, grammar, and lexical choice. Most importantly, AI systems allow simulating conversations with learners, helping them develop both accuracy and fluency in speaking. There is evidence that ChatGPT Voice is especially effective for improving the fluency, pronunciation, grammatical accuracy, lexical richness and speaking interaction with others; in particular, there is an improvement in fluency and communicative interaction skills (Hernandez-Vasquez et al., 2026).

Recently, some studies were conducted showing the capacity of voice-enabled AI chatbots to improve L² speaking. According to Hernandez-Vasquez et al. (2026), university-level EFL learners who used ChatGPT Voice achieved average growth across all dimensions of speaking skills assessed, with considerable improvements in fluency and ability to communicate with others. The participants also felt more confident about their oral performance in English, experienced less anxiety, and practiced more often. In a similar vein, ChatGPT Voice as an AI-mediated speaking companion proved to be effective in terms of five crucial criteria: fluency, confidence, vocabulary, the length of responses, and decreased speaking anxiety (Wahyuningtyas & Suparmi, 2026). Some other research shows that AI chatbots can help to develop speaking skills and reduce foreign language speaking anxiety as the low pressure created by AI assistants promotes willingness to communicate and communicative competence (Wang et al., 2024; Yunita et al., 2025).

1.2 Statement of the Problem

While voice tools based on AI become increasingly popular and advanced, the amount of empirical studies aimed at investigating their efficacy for EFL speaking improvement still is rather small. Even though the early research shows promising results connected with AI-assisted speaking training, these studies are often limited by small sample size, short time span of interventions, and single focus on specific linguistic skills. There is evidence of heterogeneous impacts on different components of oral proficiency: some of them showed considerable improvement in speaking skills while there are certain problems with feedback accuracy and possibility of technology dependency (Yan & Zou, 2024). Additionally, most of the studies related to the problem investigate AI-mediated conversations via text or AI tools

which are intended for pronunciation training; there is much less research dedicated to voice AI conversation which helps to develop speaking skills comprehensively.

Moreover, the unique features of the Ethiopian EFL context are underrepresented in the AI-related literature on language learning. Students in Ethiopian universities rarely face English speech outside classes, they have no chance to communicate with native or advanced English speakers. The class size is large and there is little focus on real-life speaking activities. In such circumstances, the benefits of AI voice tools can be examined in detail to determine whether they can be helpful in augmenting teaching and speaking practice.

In this study, the issue of generative AI interaction is explored in the Ethiopian EFL context over the period of 12 weeks. Three central elements of speaking development will be investigated: fluency, accuracy, and willingness to communicate (WTC) which is an important affective factor.

1.3 Research Questions

This study addressed the following research questions:

What are the effects of generative AI-powered voice tools on EFL learners' speaking fluency, as measured by speech rate and mean length of utterance?

What are the effects of generative AI-powered voice tools on EFL learners' speaking accuracy, as measured by grammatical and lexical accuracy?

What are the effects of generative AI-powered voice tools on EFL learners' willingness to communicate?

What are EFL learners' perceptions of using generative AI-powered voice tools for speaking practice?

1.4 Objectives of the Study

The specific objectives of this study were to:

Compare the speaking fluency development of students using generative AI voice tools with those receiving traditional speaking instruction.

Compare the speaking accuracy development of students using generative AI voice tools with those receiving traditional speaking instruction.

Examine changes in WTC among students using generative AI voice tools.

Explore students' perceptions of AI-mediated speaking practice.

1.5 Importance of the Study

The present study adds to the growing body of literature on the topic of artificial intelligence in language education through providing empirical data on an understudied English as a Foreign Language (EFL) context. Findings shed light on the potential use of generative AI voice applications to tackle the problems arising in speaking instruction in resource-scarce contexts. In line with the mission of Teaching English with Technology which involves the promotion of research “that considers human with AI approach to better teaching and learning outcomes” (Alexander & Krajka, 2025), this study provides recommendations for implementation of generative AI in EFL speaking instruction based on the findings of empirical studies.

1.6 Limitations of the Study

In particular, the study focused on 30 third-year EFL students of Jinka University. The duration of the intervention was 12 weeks, and the study assessed three dimensions of speaking proficiency development: fluency, accuracy, and willingness to communicate (WTC). Other dimensions of speaking proficiency development such as pronunciation and interactional competence were not considered in this study.

Literature Review

2. Literature Review

This study is based on a number of theoretical frameworks that provide background for speaking development via the use of artificial intelligence (AI). As formulated by Long (1996) in the Interaction Hypothesis of the Interactionist Theory of Second Language Acquisition (SLA), language acquisition takes place as a result of interaction, meaning negotiation, and modified input and output. Under this hypothesis, conversational interactions mediated by the voice functions of AI technology afford a learner the chance to be engaged in interactions which resemble those in real-life conditions, receive input tailored to the learner's level of proficiency, and produce output which is understood by the conversational agent. Based on the Interaction Hypothesis, Wang et al. (2024) studied the effects of conversational generative AI chatbots on EFL learners' willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence.

Modern researchers have described AI-mediated interaction as comprehensible input and as a medium for affective filter-free learning environments. Experimental data show that AI tools serve as dynamic sources of comprehensible input which modulate linguistic complexity, repeat exposure to lexical and grammatical patterns, and maintain interaction for an extended period. Such features correlate with the basic input characteristics found in SLA studies, such as frequency, salience, comprehensibility, and interactional relevance. Development of AI-powered conversational agents that engage L2 learners in conversations shows certain positive effects on developing speaking skills (Kanoksilapatham & Takrudkaew, 2025).

The Output Hypothesis developed by Swain (1995) focuses on the importance of language production for L2 development. In accordance with this theory, learners need chances to produce language in a meaningful way since the output forces a deeper processing, provides noticing of linguistic gaps, and helps in testing the linguistic hypotheses. Voice AI tools provide great opportunities for language production in the environment which is low-anxiety with immediate feedback. Studies have shown that learners have attributed their progress in speaking to engaging in meaningful communicative activities, receiving comprehension feedback, and practicing in non-threatening environment (Wahyuningtyas & Suparmi, 2026).

Also, the current study considers the concept of Willingness to Communicate (WTC) in relation to learners' willingness to participate in communication activities if the chance arises. According to empirical data, the formation of WTC depends on two types of factors: personal (language anxiety, perceived competence) and situational (task, interlocutor). AI-supported interaction can contribute to higher WTC by helping to overcome anxiety and creating a friendly atmosphere for oral communication practice. The research on conversational

generative AI chatbots proved that such chatbots are capable of increasing WTC due to the provision of immersive and emotional support of the learning experience (Wang et al., 2024). It was found that AI chatbots can successfully lower speech anxiety and help with pronunciation, providing dynamic language-learning experience (Yunita et al., 2025).

2.2 Conceptual Framework for AI-Supported Speaking Practice

The conceptual framework proposed in the current study brings together the interactionist approach and the specific features of generative AI voice technologies. The connections among AI-based interaction, speaking skills, and affective factors are shown in Table 1.

Table 1

Conceptual Framework for AI-Mediated Speaking Development

AI Tool Affordances	Mediating Processes	Expected Outcomes
Immediate feedback	Noticing and error correction	Speaking accuracy
Extended practice time	Automatization	Speaking fluency
Low-anxiety environment	Reduced affective filter	Increased WTC
Interactive conversation	Negotiation of meaning	Communicative competence

According to the framework, the generative AI voice tool offers three primary affordances such as instantaneous feedback, prolonged practice opportunity, and an interaction with the minimal affective filter. These affordances lead to mediating cognitive and affective processes such as noticing and error correction, automatization, the minimal affective filter, and meaning negotiation, which results in the improvement of speaking performance.

Recent studies on the effect of voice ChatGPT support the presented framework. The empirical data shows that AI chatbots improve speaking skills of the users and reduce the anxiety connected with foreign language speaking; the less stressful environment created by AI assistants positively influences WTC and self-perceived communicative competence (Wang et al., 2024). The studies related to the effect of ChatGPT Voice on WTC of non-major students' state about improvements of the learners' WTC and communicative competence and reduction of speaking anxiety (Pham et al., 2025).

The use of AI tools for speaking learning has been considered in the context of sociocultural theory and ZPD. It has been found out that the role of MKO can be played by AI conversational agent; the assistance provided by the agent helps learners to do more than they can independently do. It has been discovered that learners using AI agents for speaking practice make significant progress in self-efficacy, learning interest, learning motivation, and speaking performance.

2.3 Empirical Studies on AI and Speaking Development

2.3.1 AI and Speaking Fluency

According to the proposed framework, there are three primary affordances of generative AI voice tools, which are feedback, increased time for practice, and low-anxiety interactional context. Those affordances elicit certain cognitive and affective mediations (noticing and error correction, automatization, lower affective filter, and negotiation of meaning) that altogether lead to better speaking outcomes.

There is evidence for support of this framework in recent research of voice-enabled ChatGPT. The results of research suggest that AI chatbots greatly improve speaking proficiency and reduce foreign language speaking anxiety as well as increase willingness to communicate and self-perceived communicative competence (Wang et al., 2024). There are studies that examined how ChatGPT Voice influences the WTC of non-major students. According to the results, the use of this tool leads to positive changes in learners' WTC, increased communicative confidence, and reduction in speaking anxiety as the result of less intimidating environment (Pham et al., 2025).

The integration of AI tools into speaking instruction has been investigated from the perspective of sociocultural theory and ZPD. There is evidence that AI conversational agents can be MKOs who give learners necessary scaffolding that allows them to achieve more than they are capable of doing on their own. It was found that the learners who use AI agents for oral practice show significant gains in learning self-efficacy, learning interest, learning motivation, and speaking proficiency, and they consider such practice to be beneficial for them (Cheon, 2023).

In all AI-mediated speaking interventions, positive effects on fluency have been observed. According to the results of the systematic review done by Jeon, Lee, and Choi (2023), the use of speech-recognition chatbots in language learning gives promising improvements in fluency and oral performance. Hsu et al. (2023) showed that task-oriented chatbots provide effective practice opportunities for EFL learners leading to increased fluency and confidence.

Some recent researches focused specifically on the effects of generative AI tools. It was shown by Hernandez-Vasquez et al. (2026) that the use of ChatGPT Voice considerably increases university students' English speaking proficiency with a significant gain in fluency and communicative interaction due to the extended speaking time and repeated practice with this tool. Hernández et al. (2026) have shown that after the use of ChatGPT Voice, the most considerable improvement is achieved in fluency and coherence, with an increase in fluency of 1.3 points and general improvement of 1.0 points, as well as increased confidence in English expression and decrease in speaking anxiety.

The effects of AI-mediated interactions on speaking skills and WTC of EFL learners were studied with the use of a quasi-experimental design by Fathi, Rahimi, and Derakhshan (2024). The results of this research suggested that AI-mediated interactions led to significant improvement in speaking performance and WTC of students in comparison with a control group. This research documented improvement in fluency, lexical richness, accuracy, pronunciation, and WTC, as well as the perception of personalized feedback and reduced pressure from it.

In the comparison of voice-enabled ChatGPT and peer practice among Korean middle school students, it has been found that the group that used ChatGPT significantly exceeded the performance of the peer interaction group in terms of L2 speaking proficiency with much higher gains in vocabulary and all affective and metacognitive dimensions (Kim & Lee, 2023). The ChatGPT group also showed positive changes in reduced anxiety and WTC in the classroom. The impact of AI-powered voice tools on fluency, interactional competence, and autonomy of EFL learners was studied. It has been found that AI-mediated speaking practice improves fluency due to increased speaking time and reduced anxiety (Cheon, 2023).

Recent investigations have examined cognitive load and learning efficiency in AI-assisted speaking instruction. Feng (2025) found that AI-supported strategies significantly reduced learners' extraneous cognitive load and improved language outcomes compared to traditional CALL approaches, with adaptive feedback emerging as the most effective strategy. Learners reported that AI tools reduced stress, improved focus, and offered personalized guidance, highlighting the role of adaptive technology in managing cognitive demands during speaking practice.

2.3.2 AI and Speaking Accuracy

The research into AI and speaking accuracy has mainly focused on the feedback provided by AI systems. Empirical research has revealed that the evaluation systems based on artificial intelligence help improve pronunciation, grammar, and lexical richness. For instance, Chen et al. (2023) conducted a study of college students' perception of Google Assistant as a tool for language learning, stating that the tool was valuable because of the feedback on language accuracy.

Research related to Google Assistant has demonstrated its ability to improve learner autonomy, give instant feedback, and reduce speaking anxiety, which positively affects fluency and pronunciation of the language (Chen et al., 2023). Research on AI-assisted speaking instruction, including quasi-experimental studies, evaluates fluency and accuracy of oral performance. It has been discovered that AI tools facilitate fluency due to creating interaction-rich environments; however, accuracy requires additional form-focused instructions (Yan & Zou, 2024).

There are still some limitations related to AI tools, such as inability to follow the fast-paced conversation and difficulties in understanding the accents of the non-native speaker (Kim & Lee, 2023). It is stated in the literature that AI tools are most effective when they are used together with traditional teaching methods. Effectiveness of the combined corpus-based techniques and artificial intelligence on speaking practice revealed that participants used more positive linguistic features and less negative ones after training with corpora and AI (Chen et al., 2023).

Recent academic works have studied the potential of ChatGPT as a facilitator of pronunciation practice. According to Mompean (2024), ChatGPT could be used as a resource for pronunciation practice in order to perform production tasks, listening exercises, and explanations of minimal pairs, prosody, and accent variation. Limitations of the tool include inconsistent outputs, limited accent models, and lack of audio and visual support.

2.3.3 AI and Willingness to Communicate

In this context, the connection between AI-aided communication and Willingness To Communicate (WTC) is one of the most important issues in today's research. Fathi et al. (2024) have found out that AI-aided communication contributes significantly to the growth of the WTC in EFL learners. The research of conversational generative AI chatbots equipped with human-like avatars shows that they can enhance the WTC because of immersion in the learning process and emotional support provided (Wang et al., 2024).

It seems that there is evidence that the use of AI tools is able to decrease speaking anxiety, thus increasing WTC. AI environments usually provide low levels of affective barriers since they make learners feel safe because there is no fear of negative evaluation from others. Investigations into the impact of different conversational generative AI chatbots on EFL learners resulted in the growth of WTC, decrease in foreign language speaking anxiety, and increase in the self-perceived communicative competence (Wang et al., 2024).

Zhang, Meng, and Ma (2024) have performed a quasi-experimental study of the AI-speaking assistant Lora, and found out that foreign language enjoyment and WTC were positively changed for Chinese EFL students along with a decrease in the level of foreign language anxiety. It means that AI tools give the possibility to be in control of the learning process, thus creating favorable conditions for WTC. Similarly, Kim and Su (2024) found out positive effects of AI chatbots on students' WTC thanks to lower anxiety levels, higher confidence, and relaxation.

Waluyo and Pratiwi (2025) have produced a narrative meta-synthesis of empirical evidence on the role of AI chatbots in English learning in an Asian EFL context. They have found out that regular interaction with AI chatbots increases the WTC and communicative confidence due to the decrease in the level of speaking anxiety and nonjudgmental adaptive feedback. The learners felt more fluent and engaged during the usage of AI chatbots like Google Assistant, Alexa, and ChatGPT, but they sometimes had technical difficulties and lacked depth of conversation.

Pham, Dinh, and Tran (2025) have studied the impact of ChatGPT Voice on the WTC of non-majored students in a Vietnamese university context. They have found out that the tool increased WTC, communicative confidence, and made speaking more comfortable for students by decreasing the level of speaking anxiety. It is similar to the study conducted by Tai and Chen (2023) about Intelligent Personal Assistants (IPAs) including Google Assistant.

2.4 Research Gaps

However, despite all the positive findings described above, there are some gaps in the literature which should be taken into consideration. First, most of the current research focuses on text-based AI technology instead of voice-based conversational AI which enables speech interaction. Secondly, previous research was mostly conducted in East Asian countries, with almost no investigation of the phenomenon in African EFL environments. Moreover, most of the studies conducted are of short-term nature and investigate only separate outcomes of the process, but not holistic development of speaking. The literature shows that although use of AI chatbots greatly improves speaking skills, its influence on accuracy is not strong enough, therefore, additional form-focused training is needed (Yan & Zou, 2024). Moreover, the literature on language learning through AI technology shows that there are very few studies which investigate the influence of easily accessible tools on cognitive (speaking skills) and affective (WTC, anxiety, emotionality) outcomes in different cultures (Waluyo & Pratiwi, 2025).

3. Methodology

3.1 Research Paradigm and Design

In the present study, post-positivist research paradigm was used as research findings are inherently somewhat imperfect and there can be different perspectives to reality. Post-positivism aligns well with the research methods that combine quantitative speaking performance assessment and qualitative perception analysis.

The research design featured a quasi-experimental pretest-posttest control-group design. It was justified in light of the natural classroom setting and impossibility to randomly assign individual participants to the conditions under which they learn English. The experimental group learned speaking using generative AI voice technologies while the control group worked on their speaking skills within pairs and small groups as it was traditionally done. This research design resembles the previous quasi-experiments carried out for similar purposes (Fathi et al., 2024).

3.2 Participants

The subjects of the research were 30 third-year students studying EFL at Jinka University, Ethiopia. There were 18 females and 12 males aged between 20 and 24 in this group. All of them were native speakers of Amharic having studied English as a foreign language for a minimum of ten years. At the moment of conducting the study, they were enrolled in English Language and Literature course and took an obligatory speaking course.

Participants were divided into experimental and control groups based on the classes they belonged to. Section A ($n = 15$) made up the experimental group, while Section B ($n = 15$) made up the control group. Both groups were taught by the same instructor in order to exclude any possible influence of the teacher.

Table 2

Participant Demographic Characteristics

Characteristic	Experimental ($n=15$)	Control ($n=15$)	Total ($N=30$)
Gender			
Female	9	9	18
Male	6	6	12
Age (Mean, SD)	22.13 (1.25)	22.27 (1.33)	22.20 (1.28)
Years of English Study	10.87 (1.06)	11.00 (1.13)	10.93 (1.08)

3.3 Instruments

3.3.1 Speaking Proficiency Test

An oral proficiency test was conducted for pretest and posttest purposes. There were three oral tasks involved in this test: (1) 2-minute long self-introduction; (2) 3-minute long task of describing pictures; (3) 5-minute long discussion on a familiar topic. All oral tests were recorded and then transcribed.

Fluency in speaking was measured in terms of speech rate (syllables per minute) and mean length of utterance (average number of syllables per AS-unit). Speaking accuracy was measured in terms of grammatical errors per 100 words and lexical errors per 100 words. Two experienced raters separately rated all samples of speaking which resulted in the inter-rater reliability rates of $r = .87$ for fluency and $r = .84$ for accuracy. Such an approach to testing is in line with the

existing practice of assessing fluency and accuracy in the context of CAF framework (complexity, accuracy, fluency) in AI-speaking studies (Hernandez-Vasquez et al., 2026).

Table 3

Speaking Proficiency Test Scoring Rubric

Dimension	Measure	Operational Definition	Scoring Method
Fluency	Speech rate	Syllables per minute	Count of syllables / total speaking time (seconds) $\times$ 60
Fluency	Mean length of utterance	Mean syllables per AS-unit	Total syllables / total AS-units
Accuracy	Grammatical errors	Errors per 100 words	Count of grammatical errors / total words $\times$ 100
Accuracy	Lexical errors	Errors per 100 words	Count of lexical errors / total words $\times$ 100

3.3.2 Willingness to Communicate Questionnaire

The WTC questionnaire modified from MacIntyre et al. (2001) examines the participants' willingness to communicate in English within several different situations. The tool contains 20 questions, rated using a 5-point Likert scale (1 = almost never willing to 5 = almost always willing). This test has a Cronbach's alpha of .89, which indicates very high internal consistency. Past validation studies of the same tool in the context of communication through AI have found supporting evidence (Wang et al., 2024).

3.3.3 Perception Questionnaire

A 18-item perception questionnaire designed specifically for this study was used to examine how useful participants found the AI voice application and how the application influenced speaking confidence. The survey was distributed to the experimental group exclusively at the conclusion of the treatment. Studies in the literature that have examined ChatGPT Voice from the students' perspective have made use of similar survey instruments to find out how much impact AI voice applications have on speaking self-efficacy and confidence.

3.3.4 Semi-Structured Interview Guide

A semi-structured interview protocol was developed for analyzing participants' use of AI-based speaking practice. During the interviews, the perceptions of the participants on the AI application, its pros and cons, as well as their general experience using the technology were obtained. Qualitative methods have been widely used in studies involving the use of AI speaking to analyze perceptions of the learners (Wahyuningtyas & Suparmi, 2026; Hernandez-Vasquez et al., 2026).

3.4 Intervention: Generative AI-Powered Voice Tool

The experimental group conducted weekly speaking proficiency exercises using a generative AI voice generator (ChatGPT voice mode). This technology helped them engage in conversations on specific topics and get feedback for their oral performance as well as get answers from the AI on their questions. This intervention was done for 12 weeks, where each session lasted for one hour. Participants were asked to use this AI tool independently following

a specified course of topics, which progressed from common conversation to academic discussion.

Table 4
Weekly Speaking Topics for AI-Mediated Practice

Week	Topic	Task Type
1	Self-introduction and personal information	Conversational
2	Daily routines and habits	Conversational
3	Describing people and places	Descriptive
4	Expressing opinions on familiar topics	Opinion-sharing
5	Making requests and giving instructions	Functional
6	Telling stories and narrating events	Narrative
7	Discussing current events	Discussion
8	Debating controversial topics	Argumentative
9	Academic discussion: research topics	Academic
10	Academic discussion: presenting arguments	Academic
11	Review and consolidation	Mixed
12	Final practice and reflection	Mixed

For the control group, there were equivalent speaking practice sessions done using teacher-led paired and group activities. The topics covered and time allocated were identical for both groups, save for the difference in the interlocutor (whether it is an AI or a human). This study is aligned with previous literature that involves comparing AI communication to human communication in speaking development (Kim & Lee, 2023).

3.5 Data Collection Procedures

The data collection lasted for 14 weeks. In the first week, both the experimental and control groups took part in the speaking pre-test and completed the first WTC questionnaires. The second to thirteenth weeks comprised the intervention stage, during which the experimental group used the AI voice application and the control group performed regular speaking practice. The seventh week involved completing the mid-point WTC questionnaires for both groups. The last week included the speaking post-test, the final WTC questionnaires, the perception questionnaire for the experimental group, and semi-structured interviews with 10 students from the experimental group.

Table 5
Research Timeline

Phase	Week	Activity
Pre-test	Week 1	Speaking pretest, WTC questionnaire (Time 1)
Intervention	Weeks 2-13	Weekly speaking practice sessions
Mid-point	Week 7	WTC questionnaire (Time 2)

Post-test	Week 14	Speaking post-test, WTC questionnaire (Time 3), perception questionnaire, interviews
------------------	---------	--

3.6 Data Analysis Procedures

Quantitative analysis was done by using SPSS, version 26. Descriptive statistics were run for all measured variables. Independent samples t-test was used to compare pretest scores between groups. Changes over time within and between the groups were measured using repeated measures ANOVA. Paired samples t-test was conducted for pre-post comparisons within the groups. This approach was adopted from prior quasi-experimental studies on the effects of AI-based speaking instruction (Fathi et al., 2024).

Semi-structured interviews were subjected to qualitative analysis through thematic analysis, using the six-phase procedure suggested by Braun and Clarke (2006). Thematic analysis involved iterative coding and reading of the interview transcript in order to identify emerging themes. Thematic analysis is extensively used in AI-speaking research as a methodology (Wahyuningtyas & Suparmi, 2026).

4. Results

4.1 Pretest Comparability

Independent samples t-tests were conducted to compare the experimental and control groups on pretest measures. Results indicated no significant differences between groups on any pretest measure, confirming comparability prior to the intervention (see Table 6).

Table 6

Pretest Comparability Between Experimental and Control Groups

Variable	Experimental (n=15) M (SD)	Control (n=15) M (SD)	t	p
Fluency (speech rate)	82.47 (8.32)	81.93 (9.14)	.17	.866
Fluency (MLU)	6.82 (1.03)	6.75 (0.98)	.19	.851
Accuracy (grammatical errors)	12.47 (2.18)	12.73 (2.35)	-.32	.752
Accuracy (lexical errors)	8.93 (1.87)	9.07 (1.94)	-.20	.842
WTC	2.98 (0.47)	2.93 (0.52)	.28	.783

Note: MLU = mean length of utterance; WTC = willingness to communicate.

4.2 Speaking Fluency

Results for speaking fluency are presented in Table 7. Both groups demonstrated improvements from pretest to post-test; however, the experimental group showed significantly greater gains.

Table 7

Descriptive Statistics for Speaking Fluency

Measure	Group	Pretest (SD)	M	Post-test (SD)	M	Gain (SD)	M
---------	-------	--------------	---	----------------	---	-----------	---

Speech rate	Experimental	82.47 (8.32)	94.83 (9.16)	12.36 (4.78)
Speech rate	Control	81.93 (9.14)	87.67 (8.75)	5.74 (3.92)
MLU	Experimental	6.82 (1.03)	8.14 (1.21)	1.32 (0.67)
MLU	Control	6.75 (0.98)	7.23 (1.05)	0.48 (0.54)

Analysis of variance showed the presence of a statistically significant main effect of time on the speed of speech, $F(1, 28) = 38.62$, $p < .001$, $\eta^2 = .58$, and statistically significant interaction effect of time and group, $F(1, 28) = 16.34$, $p < .001$, $\eta^2 = .37$. Also, for mean length of utterance, there was observed a significant main effect of time, $F(1, 28) = 29.87$, $p < .001$, $\eta^2 = .52$, and interaction effect of time and group, $F(1, 28) = 12.45$, $p = .001$, $\eta^2 = .31$. The findings confirm previous studies that demonstrate the effectiveness of AI

4.3 Speaking Accuracy

Results for speaking accuracy are presented in Table 8. The experimental group demonstrated greater reductions in grammatical and lexical errors compared to the control group.

Table 8

Descriptive Statistics for Speaking Accuracy

Measure	Group	Pretest (SD)	M Post-test (SD)	M Reduction (SD)
Grammatical errors	Experimental	12.47 (2.18)	7.86 (1.72)	4.61 (1.54)
Grammatical errors	Control	12.73 (2.35)	10.34 (2.08)	2.39 (1.78)
Lexical errors	Experimental	8.93 (1.87)	5.67 (1.43)	3.26 (1.21)
Lexical errors	Control	9.07 (1.94)	7.73 (1.68)	1.34 (1.09)

Repeated measures ANOVA for grammatical errors revealed a significant main effect of time, $F(1, 28) = 45.21$, $p < .001$, $\eta^2 = .62$, and a significant interaction effect, $F(1, 28) = 8.93$, $p = .006$, $\eta^2 = .24$. For lexical errors, a significant main effect of time was found, $F(1, 28) = 38.14$, $p < .001$, $\eta^2 = .58$, along with a significant interaction effect, $F(1, 28) = 7.86$, $p = .009$, $\eta^2 = .22$. These findings are consistent with research demonstrating that AI feedback can support learners in improving language accuracy (Chen et al., 2023), though the interaction effect sizes suggest that complementary form-focused support may be beneficial.

4.4 Willingness to Communicate

WTC scores were measured at three time points: pre-test, mid-point, and post-test. Descriptive statistics are presented in Table 9.

Table 9

Descriptive Statistics for WTC Across Time Points

Group	Pretest M (SD)	Mid-point M (SD)	Post-test M (SD)
Experimental	2.98 (0.47)	3.24 (0.52)	3.68 (0.58)
Control	2.93 (0.52)	3.01 (0.48)	3.12 (0.51)

Repeated measures ANOVA revealed a significant main effect of time, $F(2, 56) = 32.45$, $p < .001$, $\eta^2 = .54$, and a significant time $\times$ group interaction, $F(2, 56) = 14.27$, $p < .001$, $\eta^2 = .34$. Post-hoc pairwise comparisons within the experimental group showed significant increases from pretest to mid-point ($p = .008$) and from mid-point to post-test ($p < .001$). Within the control group, no significant changes were observed. These findings align with research demonstrating that AI tools positively impact students' WTC mainly due to a decrease in anxiety, increased confidence, and a more relaxing learning environment (Wang et al., 2024; Pham et al., 2025).

4.5 Qualitative Findings: Learner Perceptions

The thematic analysis of semi-structured interviews with 10 participants of the experimental group identified four major themes: (1) Reduced Speaking Anxiety, (2) Perceived Fluency Development, (3) Limitations of AI Interaction, and (4) Integration Preferences.

Theme 1: Reduced Speaking Anxiety

All participants reported lower levels of anxiety related to speaking with the AI than those experienced while communicating with peers or teachers. For instance, one participant mentioned, "When I talk to the AI, I feel no pressure. I can make mistakes without feeling embarrassed" (Participant 3). Another participant said, "I practiced more because I was not afraid of being judged" (Participant 7). The findings are consistent with previous literature that indicates AI chatbots reduce speaking anxiety of learners in a foreign language (Wang et al., 2024). Moreover, other scholars proved that speaking fluency improves due to meaningful communication practice, receiving validation of comprehensibility, and working in a supportive atmosphere (Wahyuningtyas & Suparmi, 2026).

Theme 2: Perceived Fluency Development

Some participants reported the development of their speaking fluency skills while using the AI tool. For example, one participant noted, "I became faster in responding because I had to keep the conversation going" (Participant 5). Another participant remarked, "The AI forced me to speak continuously, and I became more confident in my speaking" (Participant 9). These perceptions are in line with quantitative results related to speaking fluency improvement and the literature that proves that ChatGPT Voice increases fluency and communication interactions (Hernandez-Vasquez et al., 2026). It was revealed in studies of voice-enabled chatbots that improvements of learners' skills occur because of additional speaking time and practice opportunities (Hernández et al., 2026).

Theme 3: Limitations of AI Interaction

However, participants pointed out several limitations of AI interaction for speaking practice. Most frequently mentioned problems were the lack of non-verbal communication, misinterpretation of answers, and the lack of human interaction. For instance, one participant remarked, "The AI sometimes did not understand my accent, and I had to repeat myself" (Participant 4). Another participant noted, "Speaking to a machine is not the same as speaking to a real person" (Participant 8). The findings are consistent with research that indicates that AI tools may face problems in sustaining complex conversations and understanding of non-native accents (Kim & Lee, 2023). Other literature highlights some problems with AI responses, which

may be incorrect sometimes, as well as with inconsistency of non-native accent processing by speech recognition systems (Yan & Zou, 2024).

Theme 4: Integration Preferences

The majority of participants expressed a wish to use AI tools as a complement to, not a substitute for, human interactions. For instance, one participant remarked, "I would like to have both AI practice and class speaking activities" (Participant 2). Another participant noted, "The AI helped me practice, but I still need real conversations with my classmates" (Participant 10). This finding implies that learners appreciate the unique features of both types of interaction for improving speaking skills. The conclusions are in line with previous literature, which suggests that AI tools are the most effective when used alongside traditional methods of teaching (Kim & Lee, 2023). This position corresponds to the idea of the "human with AI" paradigm that is recommended by Teaching English with Technology (Alexander & Krajka, 2025).

4.6 Perception Questionnaire Results

The perception questionnaire administered to the experimental group (n=15) assessed participants' views on the AI voice tool. Descriptive statistics are presented in Table 10.

Table

10

Perception Questionnaire Descriptive Statistics

Dimension	Items	M	SD
Perceived Usefulness	6	3.92	0.58
Ease of Use	5	4.01	0.63
Speaking Confidence	4	3.78	0.72
Satisfaction	3	3.84	0.67

The findings reveal that the perception of the AI voice application among the subjects was mostly positive for all analyzed criteria. Specifically, the highest ratings were reported in the case of the criterion "ease of use" (M = 4.01). This finding supports the results of previous studies which revealed that students are positively disposed towards AI-based content and that AI-based tools promote learners' autonomy and immediate feedback.

5. Discussion

5.1 Effects on Speaking Fluency

Moreover, the experimental group revealed significant progress in speaking fluency when compared with the control group, which is consistent with previous research on AI-based speaking instruction (Hernandez-Vasquez et al., 2026; Hernández et al., 2026). Increased speech rate (12.36 syllables per minute) and mean length of utterance (1.32 syllables) show that generative AI voice technologies are capable of promoting speaking fluency through increased practice and decreased processing limitations.

Theoretically, the observed enhancement in speaking fluency can be explained by the Output Hypothesis. The AI voice tool provided multiple chances for students to produce oral output, thereby making learners generate language constantly while interacting. As a result of constant output practice, learners are able to automatize their speech production processes, producing faster and fluent speech. Previous research reveals that ChatGPT Voice considerably enhances fluency and interaction skills (Hernandez-Vasquez et al., 2026).

Finally, the low-anxiety environment created during AI interaction may have positively affected the fluency of the students. Since learners do not worry about making mistakes or being evaluated negatively, they can pay more attention to their speech production process, thereby generating more fluent output. Previous studies reveal that AI-based environment reduces the impact of affective barriers such as fear of negative evaluation (Wang et al., 2024).

5.2 Effects on Speaking Accuracy

Participants of the experimental group revealed a higher decrease in the number of grammatical and lexical mistakes compared to those from the control group, which proves the effectiveness of AI-assisted practice in terms of gaining accuracy. Such results coincide with earlier research stating that feedback provided by AI can help learners improve language accuracy (Chen et al., 2023).

It is possible to assume that feedback mechanisms used in AI voice tool led to the development of accuracy among participants. They have stated that they were provided with instant feedback on their spoken output, which could help them notice errors and correct them. In studies dedicated to Google Assistant, researchers concluded that immediate feedback can improve fluency and pronunciation accuracy of users (Chen et al., 2023). Interactivity of the AI tool, which meant that the learner received reformulations or recasts, gave them an opportunity to produce revised output and repair errors.

However, some participants noted limitations related to feedback provided by AI, as there were cases when responses of the system were incorrect. Such an issue correlates with issues stated in the literature about AI feedback such as inability to manage complicated conversations and problems with perceiving non-native accent (Kim & Lee, 2023).

5.3 Effects on Willingness to Communicate

It is worth noting that the considerable increase in WTC among the participants in the experimental group serves as a significant discovery with pedagogical implications. Given the steady growth in WTC over the course of 12 weeks along with both immediate and long-term effects of the intervention, it is possible to make the conclusion that AI practice can positively affect learners' communicative intentions.

A number of aspects could contribute to the improvement of WTC among the participants of the study. Firstly, lower levels of anxiety associated with communication with the AI should be considered as a factor that increased WTC. In case participants feel less anxiety during AI interactions, the perception of their communicative competencies increases, and hence, their WTC becomes higher. Prior studies revealed the positive impact of AI chatbots on WTC and the decrease of foreign language speaking anxiety (Wang et al., 2024; Pham et al., 2025).

Secondly, repetitive success in communication with the help of the AI is expected to lead to the development of self-efficacy and self-confidence. In case learners have the experience of success in communication even with a computer system, the confidence in their ability to communicate effectively grows, and WTC increases. Previous research showed that the use of conversational generative AI chatbots promotes the development of presence and leads to higher WTC (Wang et al., 2024).

Finally, the possibility to transfer WTC from the interaction with the AI to human interaction is another aspect to be mentioned. It turned out that participants were willing to speak in class after using the AI because it acted as a means for bridging to human interaction. The transfer effect seems especially valuable for EFL learners who lack confidence in speaking English in the class environment.

5.4 Learner Perceptions and Pedagogical Implications

The positive perceptions expressed by participants about the AI voice tool mean that learners are open to using AI for speaking practice. The high ratings on the perceived usefulness and ease of use measures confirm that the tool was convenient and useful to the participants. This is consistent with previous research which shows that AI-generated products are viewed positively by students and AI tools help increase learner autonomy (Hernández et al., 2026).

However, the problems mentioned by participants, such as misunderstanding and the absence of human contact, show that there is a need for thoughtful implementation of AI tools into instruction. The participants' preference for combining AI practice with human interaction implies that AI tools need to be used as supplements to instruction, not as alternatives. This finding corresponds to those studies which show that AI tools produce the best results if used together with traditional teaching methods (Kim & Lee, 2023).

The pedagogical implications of these results are as follows: The use of AI voice tools in speaking practice also means that learners will get immediate feedback about some aspects of speaking accuracy which can be useful for detecting problems and making corrections. However, the AI feedback should be supplemented with the feedback of teachers because they can give more contextualized information and detect more aspects of speaking than automated systems can do (Chen et al., 2023). Furthermore, the greater opportunities for producing speech given by AI voice tools may be helpful for developing speaking fluency because learners will get more practice with less time limitations compared to class environment (Hernandez-Vasquez et al., 2026). Finally, regular practice with AI voice tools can help develop learners' willingness to communicate (WTC) because it can lower their speaking anxiety and make them better communicators (Wang et al., 2024; Pham et al., 2025).

6. Conclusion

6.1 Summary of Findings

The present study was conducted to investigate the influence of generative AI-driven voice tools on speaking fluency, accuracy, and WTC among English as a Foreign Language (EFL) students. The findings revealed that students who used AI tools in their practice had more significant gains in both speaking fluency and accuracy compared to students who performed speaking tasks using traditional methods. In addition, the use of AI tools resulted in a significant

increase in WTC, which lasted for the entire 12 weeks of intervention. Students expressed positive attitudes toward the AI voice tool by acknowledging its utility, usability, and role in fostering speaking confidence, as well as the drawbacks in terms of authenticity of interactions and technical problems.

6.2 Contributions to Knowledge

There are several ways through which this study has contributed to the existing literature on the use of artificial intelligence in language learning. To begin with, it has provided some evidence on the effectiveness of voice generators in helping learners improve their speaking skills in one of the understudied contexts, Ethiopian universities. Additionally, it builds on previous research by not only considering linguistic factors (fluency and accuracy), but also affective factors (willingness to communicate, WTC), all at once (Wang et al., 2024). In other words, this study adds to the existing literature not only quantitative data, but also some qualitative insights (Wahyuningtyas & Suparmi, 2026).

6.3 Limitations

It is important to acknowledge several limitations of this study. First, the sample size is rather limited ($n = 30$) limiting the possibility of extrapolation of results. Second, the research has been conducted at one institution only, and thus the transferability of findings is questionable. Third, the 12-week intervention period may be insufficient for the detection of long-term effects of AI-assisted practice. Fourth, the study was rather quantitative and involved only limited qualitative analysis. Finally, the study did not examine how speaking gains achieved due to the use of the intervention were retained over time.

6.4 Recommendations for Future Research

Future research can overcome these limitations in a number of ways. First, large-scale research involving larger sample sizes and multiple institutions should be carried out to improve the generalizability of results. Second, longitudinal studies investigating the long-term effect of the use of AI on the language development of students are necessary. Third, comparative studies comparing various types of AI tools (text-based, voice-based, embodied in avatars) will help in establishing the best way to design the AI tool for speaking practice (Yan & Zou, 2024). Fourth, research on the integration of the AI tools with certain pedagogical practices will facilitate the process of integration in the classroom. Finally, studies investigating the transfer of gains achieved through the use of AI in human interaction context are warranted.

6.5 Pedagogical Recommendations

Based on the results of the study, the following pedagogical recommendations can be made: Generative artificial intelligence voice tools must serve as supplementary means for speaking practice for EFL teachers. In other words, AI voice tools can provide learners with more chances to speak and help both with the development of fluency and accuracy in speaking. For instance, instructors may encourage speaking for extended periods of time in order to give students larger chances to produce oral speech and make use of the feedback system offered by AI tools to help students find and develop elements of speaking accuracy.

Speaking practice with AI assistance must be combined with interaction with people in order to create authentic communicative experiences for students. While the AI tools can give

frequent and convenient opportunities for practice, it is crucial for learners to have opportunities to interact with teachers and peers for the development of real-life communication skills. Moreover, students must be taught to analyze AI assistance and not to take it for granted. Students must be able to choose valuable feedback and understand its limitations and apply information for their benefit in order to become better speakers (Chen et al., 2023). Lastly, there must be professional development programs designed to prepare EFL instructors to work with AI tools.

7. Conclusion

The introduction of voice technology powered by generative artificial intelligence in teaching speaking English as a foreign language is undoubtedly a breakthrough in tackling the existing problems in the teaching of speaking. The current study has provided evidence of how this technology can be used to develop speaking fluency, accuracy, and willingness to communicate (WTC) in the context of EFL in Ethiopia. With the ongoing advancements in artificial intelligence technologies and their increasing accessibility, their role in helping language learners in different contexts is expected to widen (Yan & Zou, 2024). However, the conscious application of this technology based on proper pedagogy and teacher knowledge is important.

Funding Statement

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

Conflict of Interest Statement

The author declares no conflict of interest.

Ethics Approval Statement

Ethical approval was obtained from the Research Ethics Committee of Jinka University.

Author Contributions

The author conceptualized the study, designed the methodology, conducted the research, analyzed the data, and wrote the manuscript.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Generative AI and AI-Assisted Technologies

The authors declare that no generative AI or AI-assisted technology was used in the preparation of this manuscript.

References

- Suddaby, R. (2006). From the Editors: What Grounded Theory is Not. *Academy of Management Journal*, 49(4), 633–642. <https://doi.org/10.5465/amj.2006.22083020>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chen, M. R. A., & Lin, Y. H. (2023). AI CHATBOT-SUPPORTED GRAMMAR TENSE LEARNING ON EFL STUDENTS LEARNING EFFECTIVENESS. *INTED Proceedings*, 1, 1207–1216. <https://doi.org/10.21125/inted.2023.0353>

- Cheon, S. M. (2023). Educational application of ChatGPT: Its impact on Korean university students' English speaking skills. *The Journal of Linguistics Science*, 107, 469–496. <https://doi.org/10.21296/jls.2023.12.107.469>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Hernandez-Vasquez, R. M., Casas, L. I. Z. D. L., Navarro, J. R. C., Sánchez, M. I. C., & Salinas, J. A. E. (2026). Using ChatGPT voice to train English language speaking skills in university students in metropolitan Lima. *International Journal of English Language and Literature Studies*, 15(2), 1–11. <https://doi.org/10.55493/5019.v15i2.5914>
- Hernández, R. M., Pintado, P. R. C., Rojas, T. G. Q., Baca, J. E. G., Villarreal, J. L. M., Perez, J. R. G., Colchado, J. C. C., Saenz, J. A. C., Saenz, F. G. C., Calderón, E. A. B., Gonza, G. C. E. L., Garay, J. P. P., Huaco, M. J. A., & Albornoz, V. C. (2026). English Speaking Practice with ChatGPT Voice among Peruvian University Students. *World Journal of English Language*, 16(4), 96. <https://doi.org/10.5430/wjel.v16n4p96>
- Hsu, M. H., Chen, P. S., & Yu, C. S. (2021). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 31(7), 4297–4308. <https://doi.org/10.1080/10494820.2021.1960864>
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. <https://doi.org/10.1016/j.compedu.2023.104898>
- 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., Pfeffer, J., 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>
- Thủy, D. T., Da, C., & Hạnh, N. V. (2024). The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1328769>
- Liu, G., Darwin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 38(7), 1632–1660. <https://doi.org/10.1080/09588221.2024.2310288>
- Long, M. H. (1996). The Role of the Linguistic Environment in Second Language Acquisition. *Handbook of Second Language Acquisition*, 413–468. <https://doi.org/10.1016/b978-012589042-7/50015-3>
- MacIntyre, P. D., Baker, S., Clément, R., & Conrod, S. (2001). WILLINGNESS TO COMMUNICATE, SOCIAL SUPPORT, AND LANGUAGE-LEARNING ORIENTATIONS OF IMMERSION STUDENTS. *Studies in Second Language Acquisition*, 23(3), 369–388. <https://doi.org/10.1017/s0272263101003035>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G.,

- Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., 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>
- Pham, H. T. U., Dinh, T. T. L., & Tran, X. (2025). The Impact of ChatGPT Voice on Non-Majored Students' Willingness to Communicate in a Vietnamese University. *AsiaCALL Online Journal*, 16(1), 312–327. <https://doi.org/10.54855/acoj.2516116>
- Swain, M. (1995). Three functions of output in second language learning. *Medical Entomology and Zoology*, 125–144. <https://openalex.org/W207698892>
- Tai, T., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Wahyuningtyas, R. T., & Suparmi, S. (2026). Exploring the role of ChatGPT voice conversation feature in improving EFL students' speaking self-efficacy. *Celtic : A Journal of Culture, English Language Teaching, Literature and Linguistics*, 13(1), 195–212. <https://doi.org/10.22219/celtic.v13i1.44287>
- Waluyo, B., & Isti Pratiwi, D. (2025). AI chatbot-assisted English learning and willingness to communicate: A narrative meta-synthesis of evidence from Asian English as a foreign language contexts. *The JALT CALL Journal*, 21(3), 102884. <https://doi.org/10.29140/jaltcall.v21n3.102884>
- Wang, C., Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, 127, 103533. <https://doi.org/10.1016/j.system.2024.103533>
- Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies. *Smart Learning Environments*, 11(1). <https://doi.org/10.1186/s40561-024-00342-5>
- Yunita, R., Febrilyantri, C., Hayasaki, A., & Ni'am, M. I. (2025). Bridging the Affective-Linguistic Gap: A Mixed-Methods Exploration of AI-Assisted Speaking Practice and Willingness to Communicate. *Jurnal Pendidikan Progresif*, 15(4), 2760–2780. <https://doi.org/10.23960/jpp.v15i4.pp2760-2780>
- Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, 121, 103259. <https://doi.org/10.1016/j.system.2024.103259>