SMART
(Journal of English Language Teaching and Applied Linguistics)
Volume 12, No. 2, August 2026 Page. 250-268
p-ISSN: 2356-2048 e-ISSN: 2356-203x

AI-ASSISTED SPEAKING PRACTICE AND EFL STUDENTS' SPEAKING FLUENCY AND CONFIDENCE

Masitoh Oktavia[1)*], **Dalman**[2]
[1)*, 2)] University of Muhammadiyah Lampung
Email Correspondence: masitohvia@gmail.com

Abstract

This study examined the impact of AI-assisted speaking practice on EFL students' speaking fluency and confidence at Universitas Muhammadiyah Lampung. A quasi-experimental pretest–posttest control group design involved 80 third-semester English Education students. The experimental group received eight weeks of AI-assisted speaking practice using ChatGPT Voice and Gemini Live. Results indicated significant improvements in speaking fluency ($t = 7.044$, $p < 0.001$) and speaking confidence ($t = 7.710$, $p < 0.001$), with very large effect sizes ($d = 2.91$ and 3.82, respectively). These findings demonstrate the effectiveness of AI-assisted speaking practice in enhancing EFL students' fluency and confidence.

Keywords: *AI-assisted speaking, EFL, Generative AI, Speaking Fluency, Speaking Confidence*

INTRODUCTION

The ability to communicate effectively in English has become an essential competence in the twenty-first century, driven by globalization, technological advancement, and increasing international collaboration. In higher education, speaking competence is particularly important because it enables students to participate in academic discussions, professional communication, and global knowledge exchange. For students majoring in English Education, speaking proficiency is not only a learning outcome but also a professional competence required for their future careers as English teachers and educators. Consequently, improving students' speaking ability has become one of the primary goals of English as a Foreign Language (EFL) instruction, encouraging universities to adopt innovative pedagogical approaches that foster communicative competence and learner engagement (Ng et al., 2025; Nhan et al., 2026; Yassin, 2024).

Despite this growing emphasis, speaking remains one of the most challenging language

Creation is distributed under the Creative Commons License Attribution Share Alike 4.0 International Published in https://ejournal.umpri.ac.id/index.php/smart/index SMART Journal: Journal of English Language Teaching and Applied Linguistics

skills for Indonesian EFL learners. Many university students experience difficulties expressing ideas fluently due to limited vocabulary, inaccurate pronunciation, inadequate grammatical control, and insufficient opportunities for authentic communication. These linguistic challenges are often accompanied by affective barriers, including speaking anxiety, fear of making mistakes, and low confidence, which may reduce students' willingness to communicate and participation in English-speaking activities (Irawan et al., 2026; Pilu et al., 2025; Yassi et al., 2023). In conventional speaking classes, opportunities for extended individual practice and immediate, individualized interaction may also be constrained by limited instructional time, class size, and teacher availability. These conditions indicate the need for additional speaking practice that allows learners to produce oral language repeatedly in a supportive environment.

Among the dimensions of oral proficiency, speaking fluency is widely regarded as a key indicator of communicative success because it reflects learners' ability to produce coherent and spontaneous speech with minimal hesitation. However, speaking performance is influenced not only by linguistic competence but also by psychological factors. Speaking confidence, which is closely related to learners' self-efficacy, determines their willingness to initiate conversations, express opinions, and actively participate in communicative activities. Students with higher confidence generally demonstrate greater persistence, engagement, and speaking performance than those with lower confidence. Therefore, effective speaking instruction should simultaneously improve students' linguistic competence and confidence to communicate (Tsang & Lee, 2023; Zhang et al., 2026).

Recent advances in Artificial Intelligence (AI), particularly generative AI technologies such as ChatGPT, Gemini, and Microsoft Copilot, have created new possibilities for language learning by providing interactive and personalized learning environments. Unlike traditional digital learning resources, these AI-powered systems allow learners to engage in real-time conversations, receive immediate conversational responses, and practice speaking repeatedly without being constrained by classroom schedules or the availability of human interlocutors. For speaking instruction, voice-enabled systems are particularly relevant because they allow learners to practice oral production through direct spoken interaction rather than relying primarily on text-based prompts and responses. ChatGPT Voice and Gemini Live therefore

offer potential opportunities for repeated conversational practice that can complement teacher-led classroom activities. Such characteristics make generative AI a promising instructional tool for developing oral communication skills and supporting learner autonomy. (Z. Chen et al., 2026; Cheon et al., 2025; Goh & Aryadoust, 2025).

The increasing adoption of AI in language education has generated substantial research interest. Previous studies have reported positive contributions of AI-supported learning to various aspects of language learning, including writing quality, vocabulary acquisition, grammar accuracy, pronunciation, learner engagement, and autonomous learning (Bahari, 2026; J. Chen, 2024; Syafei, 2025). Other studies indicate that AI-powered conversational systems provide additional opportunities for communication practice and may support learners' affective engagement in English learning. Overall, the existing literature suggests that AI has considerable pedagogical potential for supporting language learning. However, the evidence remains fragmented because most studies have focused on isolated language skills or learners' perceptions rather than providing comprehensive evidence regarding measurable development in oral communication outcomes, particularly fluency and confidence examined together.

Nevertheless, several important research gaps remain. First, existing studies have predominantly examined AI as a writing assistant or a tool for vocabulary and grammar learning, whereas relatively limited attention has been given to its role as an interactive speaking partner. Second, research on AI-assisted speaking has generally investigated either speaking performance or learners' perceptions separately, with few studies examining speaking fluency and speaking confidence simultaneously. Third, rigorous quasi-experimental studies comparing AI-assisted speaking practice with conventional classroom instruction remain scarce, particularly in the context of Indonesian private universities. More specifically, there is limited empirical evidence showing whether repeated voice-based interaction with generative AI produces greater improvements in speaking fluency and confidence than conventional speaking instruction. This comparison is important because the pedagogical value of AI should be demonstrated through measurable learning outcomes rather than through its technological novelty or students' positive perceptions alone.

The novelty of this study lies in its direct comparison of voice-based AI-assisted speaking practice and conventional speaking instruction while simultaneously measuring both speaking fluency and speaking confidence. Specifically, ChatGPT Voice and Gemini Live are used as interactive speaking partners that provide students with repeated opportunities to respond orally, sustain conversations, and practice English without depending on continuous teacher or peer availability. In contrast, conventional speaking instruction relies primarily on teacher-led activities, classroom interaction, and scheduled practice opportunities. Comparing these two approaches enables the study to determine whether the additional affordances of voice-based generative AI, particularly repeated individual oral practice and immediate conversational interaction, result in measurable improvements in both linguistic and affective outcomes. Thus, rather than merely examining whether students use or perceive AI positively, this study provides quasi-experimental evidence of whether AI-assisted speaking practice is more effective than conventional instruction in developing EFL students' speaking fluency and confidence.

Addressing these gaps, the present study investigates the effectiveness of AI-assisted speaking practice using ChatGPT Voice and Gemini Live in improving speaking fluency and speaking confidence among undergraduate students of the English Education Study Program at Universitas Muhammadiyah Lampung. By employing a quasi-experimental design with a conventional-instruction comparison group, the study provides empirical evidence concerning the effectiveness of voice-based generative AI for two complementary dimensions of speaking development: linguistic performance and affective readiness to communicate. The findings are expected to contribute to the growing literature on AI-assisted language learning while offering practical implications for English lecturers, curriculum developers, and higher education institutions seeking to integrate voice-based generative AI into speaking instruction.

Accordingly, this study addresses the following research objectives: (1) to investigate the impact of AI-assisted speaking practice on EFL students' speaking fluency and (2) to examine its impact on students' speaking confidence. Based on these objectives, two hypotheses were formulated: H1: AI-assisted speaking practice significantly improves speaking fluency; and H2: AI-assisted speaking practice significantly improves speaking

confidence. The study is expected to provide empirical evidence supporting the effective integration of generative AI into English speaking instruction in higher education.

RESEARCH METHOD

This study employed a quantitative approach using a quasi-experimental research design with a pretest–posttest control group design. The design was selected because the participants were assigned to existing intact classes rather than individually randomized into experimental and control groups. Two third-semester classes of the English Education Study Program at Universitas Muhammadiyah Lampung participated in the study. The two classes were selected from the available third-semester classes based on their comparable academic level and the feasibility of implementing the intervention within the existing class structure. One intact class was designated as the experimental group, while the other was designated as the control group. The assignment was made at the class level rather than at the individual student level to preserve the existing classroom organization. The experimental group received AI-assisted speaking practice, whereas the control group received conventional speaking instruction. Both groups completed a pre-test before the intervention and a post-test after the intervention to evaluate changes in speaking fluency and speaking confidence.

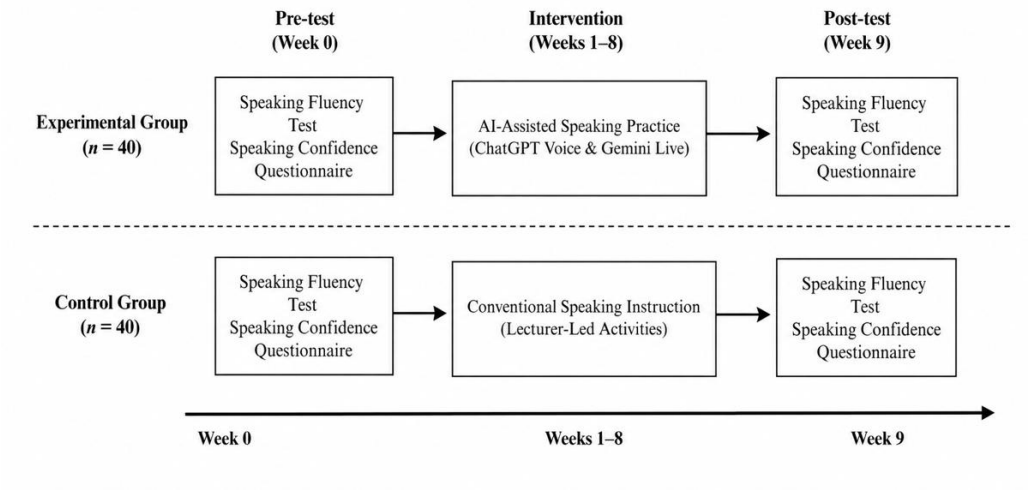


Figure 1. Research Design

The study was conducted at Universitas Muhammadiyah Lampung and involved undergraduate students enrolled in the English Education Study Program. The participants

consisted exclusively of third-semester students from two existing classes. No participants from other institutions or educational levels were included in the study. The two classes were selected because they represented the same semester and study program and followed comparable speaking-related learning objectives. One class was assigned as the experimental group and the other as the control group based on the existing class arrangement and the feasibility of implementing the intervention. This procedure was consistent with the use of intact classes in a quasi-experimental design.

The participants consisted of 80 third-semester undergraduate students enrolled in the English Education Study Program at Universitas Muhammadiyah Lampung. Forty students were assigned to the experimental group and forty to the control group. The experimental and control groups were formed from two pre-existing classes rather than through individual randomization. The groups followed the same institutional curriculum, studied comparable speaking topics and learning objectives, and received instruction during the same academic period. The study was conducted during one academic semester over an eight-week period. The experimental and control groups studied the same speaking topics and learning objectives; however, they differed in the instructional approach employed.

Table 1. Demographic Characteristics of Participants

Characteristics	Experimental	Control	Total
Number of Students	40	40	80
Semester	Third	Third	–
Study Program	English Education	English Education	–
Institution	Universitas Muhammadiyah Lampung	Universitas Muhammadiyah Lampung	–
Duration	8 Weeks	8 Weeks	–

The intervention lasted for eight weeks with two instructional meetings per week, resulting in sixteen learning sessions. Students in the experimental group practiced speaking using ChatGPT Voice and Gemini Live as AI conversational partners. The learning activities included guided conversations, role plays, interviews, storytelling, discussions, and opinion-sharing tasks. The AI platforms were used specifically to provide repeated opportunities for oral interaction, allowing students to respond to prompts, sustain conversations, and practice

spoken English during the designated speaking activities. Throughout the intervention, AI provided immediate conversational responses and opportunities for repeated speaking practice. In contrast, the control group received conventional lecturer-led speaking instruction using classroom discussions, presentations, and pair-work activities without AI assistance. Both groups received the same speaking topics, instructional objectives, and equivalent intervention duration; the primary difference was the use of voice-based AI as an additional speaking partner in the experimental group.

Table 2. AI-Assisted Speaking Activities

Week	Speaking Topic	AI Platform	Main Activity
1	Self-Introduction	ChatGPT Voice	Conversation
2	Daily Activities	Gemini Live	Dialogue
3	Education	ChatGPT Voice	Discussion
4	Tourism	Gemini Live	Role Play
5	Social Issues	ChatGPT Voice	Interview
6	Technology	Gemini Live	Presentation
7	Culture	Gemini Live	Storytelling
8	Reflection	ChatGPT Voice	Discussion

Two research instruments were employed to collect the data: a speaking performance test and a speaking confidence questionnaire. The speaking performance test was administered as a pre-test and post-test to assess students' oral performance. An analytical scoring rubric comprising five components, fluency, pronunciation, grammar, vocabulary, and comprehension, was used to evaluate the speaking performance. Because the primary linguistic outcome of this study was speaking fluency, the fluency component was used as the principal measure of speaking fluency, while the other components provided complementary information about overall oral performance.

Speaking confidence was measured using a 20-item questionnaire administered before and after the intervention. Each item was rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire measured students' perceived confidence in performing speaking activities, willingness to communicate, perceived ability to speak English, and motivation to participate in oral communication. Items referring directly to anxiety reduction were included only if they were part of the validated instrument and were interpreted as indicators of students' speaking-related affective experience rather than as a separate outcome variable.

Table 3. Research Instruments

Instrument	Variable	Indicators
Speaking Test	Speaking Fluency	Fluency, Pronunciation, Grammar, Vocabulary, Comprehension
Questionnaire	Speaking Confidence	Confidence, Willingness to Communicate, Anxiety Reduction, Speaking Motivation

Table 4. Speaking Assessment Rubric

Assessment Component	Score
Fluency	1–5
Pronunciation	1–5
Grammar	1–5
Vocabulary	1–5
Comprehension	1–5

Before implementation, the research instruments were evaluated to establish their validity and reliability. The speaking confidence questionnaire underwent content validation through expert judgment by three experts with expertise in English language teaching and educational assessment. The content validity result was CVI = 0.95 (Content Validity Index), indicating that the questionnaire items adequately represented the construct of speaking confidence according to the predetermined validity criteria.

The speaking confidence questionnaire was then pilot-tested with 30 students who were not included in the main study. Its internal consistency was examined using Cronbach's Alpha, resulting in a coefficient of 0.913. This value exceeded the predetermined reliability criterion of 0.70, indicating excellent internal consistency of the questionnaire.

For the speaking performance test, the content validity of the analytical rubric was established through expert judgment by three experts, yielding a CVI of 1.00. The speaking performances were independently assessed by two raters who were lecturers with relevant expertise in English language teaching and speaking assessment. Before the final scoring, both raters used the same analytical rubric and assessment criteria to ensure consistency in the assessment process. Inter-rater reliability was examined using the Intraclass Correlation Coefficient (ICC), which yielded a coefficient of 0.983, indicating excellent agreement between the two raters.

These procedures ensured that the research instruments demonstrated adequate content validity and reliability and that the speaking performance scores were assessed consistently across raters.

Data collection began with the administration of a pre-test to both groups to determine students' initial speaking fluency and speaking confidence. The speaking performance test and speaking confidence questionnaire were administered to both groups before the intervention. Following the pre-test, the experimental group participated in AI-assisted speaking practice for eight weeks, while the control group received conventional speaking instruction covering the same learning objectives. At the end of the intervention, both groups completed the post-test using the same measurement procedures as the pre-test, including the speaking performance assessment and speaking confidence questionnaire. The resulting data were then prepared for statistical analysis.

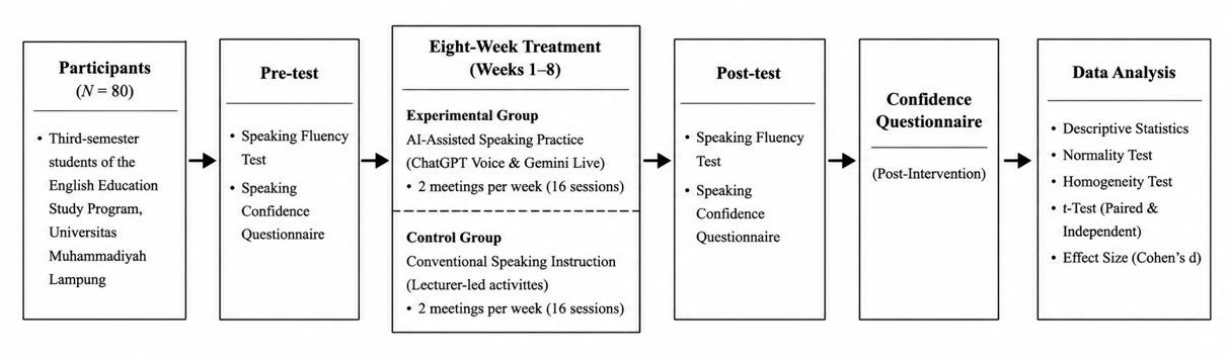


Figure 2. Research Procedures

The collected data were analyzed using IBM SPSS Statistics. Descriptive statistics were computed to summarize students' speaking performance and confidence scores. The assumptions of normality and homogeneity were examined using the Shapiro–Wilk and Levene's tests, respectively. Paired-sample t-tests were conducted to determine within-group improvements between the pre-test and post-test scores, whereas independent-sample t-tests were performed to compare post-test performance between the experimental and control groups. Because the study used intact classes rather than individually randomized participants, the comparison of post-test outcomes was interpreted with consideration of the quasi-experimental design and the baseline pre-test results. Finally, Cohen's d was calculated to determine the practical magnitude of the treatment effect.

FINDING AND DISCUSSION

Findings

Pre-test and Post-test Descriptive Statistics

Prior to the intervention, a baseline analysis was conducted to determine whether the experimental and control groups had comparable levels of speaking fluency and speaking confidence. Establishing group equivalence before the treatment is essential to ensure that any observed differences in the post-test are attributable to the intervention rather than pre-existing differences between the groups.

Table 5. Baseline Statistics of the Experimental and Control Groups

Variable	Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Gain
Speaking Fluency	Experimental	40	11.45 (1.60)	16.43 (1.95)	4.99
Speaking Fluency	Control	40	11.32 (1.76)	13.33 (1.99)	2.01
Speaking Confidence	Experimental	40	54.98 (6.64)	77.74 (8.22)	22.75
Speaking Confidence	Control	40	55.96 (7.10)	63.68 (8.09)	7.71

As shown in Table 5, the two groups had comparable pre-test means for speaking fluency (11.45 for the experimental group and 11.32 for the control group) and speaking confidence (54.98 and 55.96, respectively). Following the eight-week intervention, both groups showed improvements in the two outcome variables. However, the experimental group demonstrated larger mean gains than the control group. For speaking fluency, the experimental group increased by 4.99 points compared with 2.01 points in the control group. Similarly, speaking confidence increased by 22.75 points in the experimental group compared with 7.71 points in the control group.

These descriptive results indicate that the experimental group experienced greater numerical improvement than the control group. However, differences in mean gain alone do not establish statistical significance; therefore, inferential tests were conducted to determine whether the observed differences between the groups were statistically significant.

Improvement in Speaking Fluency

Speaking fluency was assessed before and after the eight-week intervention. The descriptive results in Table 5 indicate that the experimental group obtained a higher post-test mean (M = 16.43, SD = 1.95) than the control group (M = 13.33, SD = 1.99). The

experimental group also demonstrated a larger mean gain (4.99) than the control group (2.01).

Table 6. Descriptive Statistics of Speaking Fluency

Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Gain
Experimental	40	11.45 (1.60)	16.43 (1.95)	4.99
Control	40	11.32 (1.76)	13.33 (1.99)	2.01

The results show that both groups improved their speaking fluency during the study. However, the mean gain was more than twice as high in the experimental group. This descriptive difference suggests a stronger improvement among students who participated in AI-assisted speaking practice. The statistical significance of this between-group difference was subsequently examined using an independent-sample t-test, while the paired-sample t-test was used to determine whether each group improved significantly from pre-test to post-test.

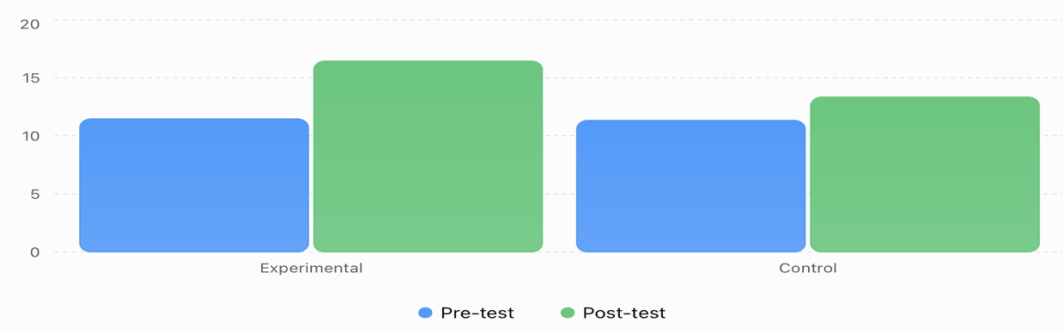


Figure 3. Improvement in Speaking Fluency

Improvement in Speaking Confidence

In addition to speaking performance, the study examined changes in students' speaking confidence before and after the intervention.

Table 7. Descriptive Statistics of Speaking Confidence

Group	N	Pre-test Mean (SD)	Post-test Mean (SD)	Mean Gain
Experimental	40	54.98 (6.64)	77.74 (8.22)	22.75
Control	40	55.96 (7.10)	63.68 (8.09)	7.71

As presented in Table 7, the experimental group showed a substantially larger increase in speaking confidence than the control group. The experimental group's mean confidence score increased by 22.75 points, whereas the control group's score increased by 7.71 points. These descriptive results indicate a considerably greater improvement in the experimental

group; however, the conclusion regarding treatment effectiveness was based on inferential comparisons rather than on the magnitude of the t-values alone.

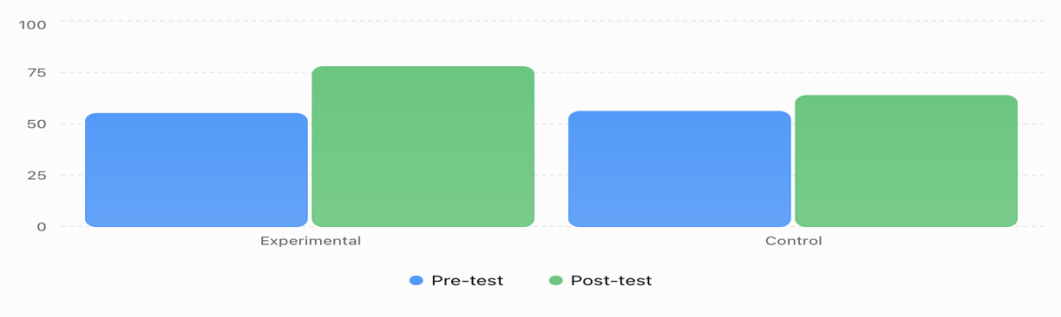


Figure 4. Improvement in Speaking Confidence

The figure demonstrates a larger increase in speaking confidence among students in the experimental group than among those in the control group.

Assumption Testing

Prior to hypothesis testing, the assumptions for parametric statistical analysis were examined using the Shapiro–Wilk test for normality and Levene's test for homogeneity of variance. These preliminary analyses were conducted to ensure that the data met the assumptions required for parametric statistical procedures. A significance value (*p*) greater than 0.05 indicates that the data are normally distributed and that the variances between groups are homogeneous.

Table 8. Test of Normality (Shapiro–Wilk)

Variable	Group	W	p-value	Decision
Speaking Fluency (Pre-test)	Experimental	0.9824	0.7763	Normal
Speaking Fluency (Pre-test)	Control	0.9671	0.2900	Normal
Speaking Fluency (Post-test)	Experimental	0.9768	0.5733	Normal
Speaking Fluency (Post-test)	Control	0.9703	0.3685	Normal
Speaking Confidence (Pre-test)	Experimental	0.9814	0.7414	Normal
Speaking Confidence (Pre-test)	Control	0.9834	0.8130	Normal
Speaking Confidence (Post-test)	Experimental	0.9525	0.0921	Normal
Speaking Confidence (Post-test)	Control	0.9708	0.3817	Normal

The results of the Shapiro–Wilk test indicate that all significance values exceeded 0.05, demonstrating that the distributions of speaking fluency and speaking confidence scores were normal in both the experimental and control groups. Therefore, the assumption of normality was satisfied.

Table 9. Test of Homogeneity (Levene's Test)

Variable	F	p-value	Decision
Speaking Fluency (Post-test)	0.017	0.897	Homogeneous
Speaking Confidence (Post-test)	0.000	0.990	Homogeneous

The Levene's test results (Table 9) confirmed that the variances of speaking fluency (p = 0.897) and speaking confidence (p = 0.990) were homogeneous, as both significance values exceeded 0.05. Together with the Shapiro–Wilk results, these findings indicate that all assumptions for parametric statistical analysis were satisfied. Therefore, paired-sample *t*-tests and independent-sample *t*-tests were performed for hypothesis testing.

Hypothesis Testing

Because the study used a quasi-experimental pretest–posttest design with intact classes, the statistical analysis was conducted in two stages. First, paired-sample t-tests were used to examine whether each group showed significant changes from pre-test to post-test. Second, independent-sample t-tests were used to examine whether the experimental and control groups differed significantly in their post-test outcomes. The descriptive mean gains were also considered when interpreting the magnitude and direction of improvement. This analytical approach allowed the study to distinguish within-group improvement from the between-group treatment effect.

Table 10. Paired-Sample *t*-Test Results

Group	Variable	*t*	*p*	Decision
Experimental	Speaking Fluency	18.42	< 0.001	Significant
Experimental	Speaking Confidence	24.18	< 0.001	Significant
Control	Speaking Fluency	9.15	< 0.001	Significant
Control	Speaking Confidence	8.64	< 0.001	Significant

As presented in Table 10, both groups demonstrated statistically significant improvements in speaking fluency and speaking confidence from pre-test to post-test ($p < 0.001$). The experimental group showed significant improvement in speaking fluency ($t = 18.42$, $p < 0.001$) and speaking confidence ($t = 24.18$, $p < 0.001$). The control group also demonstrated significant improvement in speaking fluency ($t = 9.15$, $p < 0.001$) and speaking confidence ($t = 8.64$, $p < 0.001$).

The paired-sample results therefore demonstrate improvement within both groups but do not, by themselves, establish that AI-assisted speaking practice was more effective than conventional instruction. The between-group comparison was therefore examined using independent-sample t-tests.

Table 11. Independent-Sample *t*-Test Results

Variable	*t*	df	*p*	Decision
Speaking Fluency	7.044	78	< 0.001	H1 Accepted
Speaking Confidence	7.710	78	< 0.001	H2 Accepted

The independent-sample t-test results revealed statistically significant differences between the experimental and control groups in both outcome variables. For speaking fluency, the experimental group achieved a significantly higher post-test score than the control group ($t = 7.044$, $df = 78$, $p < 0.001$). The same pattern was observed for speaking confidence ($t = 7.710$, $df = 78$, $p < 0.001$). Combined with the substantially larger mean gains observed in Table 5, these findings provide evidence that students receiving AI-assisted speaking practice achieved better post-test outcomes than students receiving conventional speaking instruction. Thus, H1 and H2 were supported.

Effect Size Analysis

Statistical significance does not necessarily indicate the practical magnitude of an educational intervention. Therefore, Cohen's d was calculated as a between-group effect size using the difference between the experimental and control group post-test means divided by the pooled standard deviation. The pooled standard deviation was calculated as follows:

$$d = (M_1 - M_2) / SD_p$$

where SD_p represents the pooled standard deviation of the two groups. This calculation was selected because the primary purpose of the effect-size analysis was to estimate the magnitude of the difference between the experimental and control groups after the intervention.

Table 12. Between-Group Effect Size of AI-Assisted Speaking Practice

Variable	Experimental Post-test M (SD)	Control Post-test M (SD)	Cohen's d	Interpretation
Speaking Fluency	16.43 (1.95)	13.33 (1.99)	1.57	Very Large
Speaking Confidence	77.74 (8.22)	63.68 (8.09)	1.72	Very Large

The effect-size analysis indicated that the differences between the experimental and control groups were not only statistically significant but also substantial in practical magnitude. The between-group effect size for speaking fluency was d = 1.57, while the effect size for speaking confidence was d = 1.72. Both values substantially exceeded the conventional threshold of 0.80 for a large effect. These results indicate that the experimental group performed considerably better than the control group following the intervention, with the largest between-group effect observed for speaking confidence.

Importantly, the effect sizes reported here differ from the previously reported values of d = 2.91 for speaking fluency and d = 3.82 for speaking confidence. Those values should not be presented as the primary treatment effect unless they are explicitly identified as within-group effect sizes and their calculation method is reported. For the present quasi-experimental comparison, the between-group effect sizes provide a more direct estimate of the relative effectiveness of AI-assisted speaking practice compared with conventional instruction.

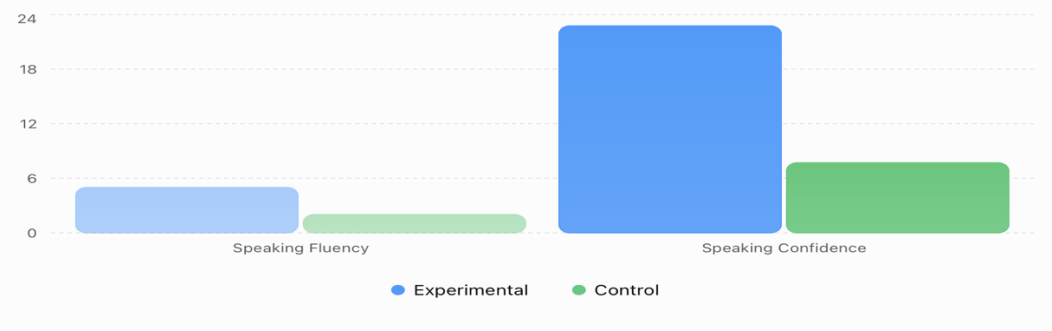


Figure 5. Comparison of Gain Scores between the Experimental and Control Groups

Discussion

The findings demonstrate that AI-assisted speaking practice was associated with significantly greater improvement in EFL students' speaking fluency than conventional speaking instruction. The experimental group achieved higher post-test scores and greater mean gains than the control group. This finding can be explained through interaction-based language learning, in which repeated opportunities to produce language, respond to conversational input, and maintain interaction may support greater fluency. Voice-based AI provided additional opportunities for oral production and repeated practice, which may have

helped students develop greater continuity and reduce hesitation during speaking. These findings are consistent with previous studies reporting positive effects of AI-powered conversational tools on speaking performance and communication opportunities (Mavidi, 2025; Taeza, 2025). Similarly, recent studies have shown that AI-assisted language learning enhances students' communicative competence, pronunciation, vocabulary use, and speaking fluency by providing interactive dialogue and immediate responses that resemble authentic human conversations (Zou & Wang, 2024). Therefore, the findings of this study extend previous evidence by confirming that AI-assisted speaking practice is also effective in improving speaking fluency among Indonesian EFL university students.

The study also found significantly greater improvement in speaking confidence among students in the experimental group. From a self-efficacy perspective, repeated successful speaking experiences may strengthen learners' beliefs in their ability to perform speaking tasks. Interaction with AI may also provide a less socially evaluative environment in which students can rehearse responses and repeat tasks. However, because this study measured speaking confidence rather than anxiety directly, reduced anxiety should be regarded as a possible explanation rather than a demonstrated mechanism. These findings are also consistent with previous studies indicating that AI-based language learning reduces speaking anxiety while promoting learners' willingness to communicate and participate more actively in classroom interactions (Du, 2026; Kang, 2025; Wang et al., 2024).

The improvement in the control group is also important. Students receiving conventional instruction showed significant gains because they continued to participate in discussions, presentations, and pair-work activities throughout the eight-week intervention. Thus, the results do not indicate that improvement was caused exclusively by AI. Rather, the larger gains in the experimental group suggest an additional benefit associated with AI-assisted practice. Increased practice time, greater opportunities for individual oral production, and repeated interaction may partly explain this difference. Nevertheless, familiarity with the pre-test and post-test tasks and the novelty of using AI may also have contributed to the observed improvement.

The findings have practical implications for higher education. AI-assisted speaking practice can be used as a complementary strategy rather than a replacement for lecturers,

providing additional opportunities for oral practice through conversations, role plays, storytelling, and discussions. However, several limitations should be considered. AI use depends on reliable devices, internet access, and students' digital competence; AI-generated feedback may not always be accurate; and students may differ in the intensity and frequency of AI use. In addition, the eight-week intervention may have allowed a novelty effect to influence engagement. Therefore, lecturers should monitor AI use, verify AI feedback, and maintain opportunities for meaningful human interaction.

Overall, this study provides evidence that integrating voice-based generative AI into speaking instruction can produce greater improvements in speaking fluency and confidence than conventional instruction alone. The contribution of the study is therefore best understood as evidence of an additional pedagogical benefit of AI-assisted practice, rather than evidence that AI independently accounts for all students' improvement.

CONCLUSION

This study found that AI-assisted speaking practice using ChatGPT Voice and Gemini Live significantly improved speaking fluency and confidence among third-semester EFL students at Universitas Muhammadiyah Lampung compared with conventional instruction. AI can therefore serve as a complementary tool to provide additional speaking practice and interaction. However, the findings are limited to 80 students at one institution over eight weeks. Future studies should involve larger samples, longer interventions, and examine AI usage intensity and feedback accuracy.

REFERENCES

Bahari, A. (2026). Evaluating the Effectiveness of AI-Driven Approaches on EFL Learners' Expository Writing Skills. *TESOL Journal*, *17*(1). https://doi.org/10.1002/tesj.70099

Chen, J. (2024). AI-Based Feedback Model in Supporting Mobile-Assisted Language Learning Environment. *2024 1st International Conference on Software, Systems and Information Technology, SSITCON 2024*. https://doi.org/10.1109/SSITCON62437.2024.10796990

Chen, Z., Wei, W., & Zou, D. (2026). Generative AI technology and language learning: global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, *34*(2), 907–920. https://doi.org/10.1080/10494820.2025.2511248

Cheon, G., Choi, Y., Lee, D., & Baek, J. (2025). Generative AI Agents in Language Learning: A Randomized Field Experiment. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 4949–4958. https://doi.org/10.24251/hicss.2025.596

Du, J. (2026). Effects of Generative AI Chatbot on English Learners' Speaking Anxiety: A Case Study. *Lecture Notes on Data Engineering and Communications Technologies*, *279*, 221–230. https://doi.org/10.1007/978-981-95-4423-3_14

Goh, C. C. M., & Aryadoust, V. (2025). Developing and assessing second language listening and speaking: Does AI make it better? *Annual Review of Applied Linguistics*, *45*, 179–199. https://doi.org/10.1017/S0267190525100111

Irawan, L. A., Sotlikova, R., & Alemayehu, M. (2026). Technological and digital literacy challenges in implementing flipped learning: insights from Eastern Indonesia. *International Journal of Evaluation and Research in Education*, *15*(2), 1776–1786. https://doi.org/10.11591/ijere.v15i2.37784

Kang, E. Y. (2025). Enhancing L2 Learners' Affective Outcomes and Oral Proficiency Through AI-Chatbot Interaction. *Korean Journal of English Language and Linguistics*, *25*, 1299–1314. https://doi.org/10.15738/kjell.25..202509.1299

Mavidi, P. N. (2025). Chatbots for conversation: AI's role in transforming English learning. In *Application of AI in the Teaching and Learning of English as a Foreign Language (EFL)* (pp. 179–222). https://doi.org/10.4018/979-8-3373-0969-9.ch006

Ng, M. L., Rahman, Md. M., Rahman, M. H., Lai, T. T., & Tan, R. S.-E. (2025). Exploring Obstacles and Effectiveness of Task-Based Language Teaching (TBLT) Approach in EFL Speaking Instruction: Insights from an Adult EFL Classroom. *World Journal of English Language*, *15*(6), 151–162. https://doi.org/10.5430/wjel.v15n6p151

Nhan, L. K., Quang, L. V. N., & Hoa, N. T. M. (2026). Implementing innovative pedagogical strategies for teaching english in elementary EFL classrooms. *Multidisciplinary Science Journal*, *8*(1). https://doi.org/10.31893/multiscience.2026114

Pilu, R., Jabu, B., & Sulaiman, I. (2025). Flipped learning and its challenges: understanding students' struggles in Indonesian EFL contexts from teachers' viewpoint. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1575385

Syafei, M. (2025). Leveraging Artificial Intelligence in Writing: ELT Students' Perspectives and Experiences. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 416–430. https://doi.org/10.26803/ijlter.24.5.22

Taeza, J. (2025). The role of AI-powered chatbots in enhancing second language acquisition: An empirical investigation of conversational AI assistants. *Edelweiss Applied Science and Technology*, *9*(3), 2616–2629. https://doi.org/10.55214/25768484.v9i3.5853

Tsang, A., & Lee, J. S. (2023). The making of proficient young FL speakers: The role of emotions, speaking motivation, and spoken input beyond the classroom. *System*, *115*. https://doi.org/10.1016/j.system.2023.103047

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*. https://doi.org/10.1016/j.system.2024.103533

Yassi, A. H., Hanafiah, W., Sahib, H., Aswad, M., Nurchalis, N. F., & Azizi, Z. (2023). Exploring the Effects of Pair-Interaction Model on Improving Indonesian Adult Learners' English Proficiency. *Studies in English Language and Education*, *10*(1), 117–136. https://doi.org/10.24815/siele.v10i1.27406

Yassin, B. (2024). Enhancing English as a Foreign Language (EFL) learning in Saudi Arabia: the academic contribution of YouTube in EFL learning and cultural awareness. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1451504

Zhang, H., Wang, Y., & Jiang, Y. (2026). Oral Fluency Characteristics of Chinese English-as-a-Foreign-Language Learners Across Different Educational Stages. *Language Teaching Research*. https://doi.org/10.1177/13621688261447957

Zou, B., & Wang, C. (2024). Using an Artificial Intelligent Speaking Assessment Platform—EAP Talk—to Develop EFL Speaking Skills. In *English for Academic Purposes in the EMI Context in Asia: XJTLU Impact* (pp. 287–300). https://doi.org/10.1007/978-3-031-63638-7_12