ENHANCING EFL SPEAKING COMPETENCE THROUGH AI-SUPPORTED INSTRUCTION: A QUASI-EXPERIMENTAL STUDY AT A VIETNAMESE UNIVERSITY

NGUYEN HONG GIANG
Vietnam National University, Hanoi

Received: 10 November 2025; Revised: 20 November 2025; Accepted for publication: 29 November 2025.

Abstract

Developing speaking competence remains a major challenge in English as a Foreign Language (EFL) instruction at high level. Recent advances in artificial intelligence (AI) offer new possibilities for enhancing speaking instruction through increased practice, interaction, and feedback. This study investigates the effectiveness of AI-supported speaking instruction among undergraduate students at a university in Hanoi. Using a quasi-experimental design, the study involved 70 undergraduates, with 35 assigned to an experimental group and 35 to a control group. The experimental group received AI-assisted speaking activities (chatbots and speech recognition tools) integrated into regular instruction (20-30 minutes per session) in eight weeks, while the control group followed traditional speaking instruction. Speaking performance was assessed through pre- and post-tests and analyzed across fluency, accuracy, complexity, and interactional competence. Independent-samples and paired-samples t-tests were employed for data analysis. The results reveal no significant difference between the two groups at pre-test. However, the experimental group significantly outperformed the control group at post-test, with a large effect size. Notable gains were observed in fluency and interactional competence, significant improvements in accuracy, and more modest gains in complexity. These findings suggest that AI-supported instruction can effectively enhance EFL speaking competence in higher education.

Keywords: *Artificial intelligence in language education, EFL speaking instruction, speaking competence, higher education, Vietnamese university context.*

1. Introduction

In recent decades, the development of learners' speaking competence has been widely acknowledged as a central objective of foreign language education, particularly in English as a Foreign Language (EFL) contexts. Within higher education, speaking ability is increasingly viewed not only as a linguistic skill but also as a key indicator of graduates' communicative competence, academic readiness, and employability. However, despite its recognized importance, speaking remains one of the most challenging skills to teach and assess effectively.

In Vietnamese universities, this challenge is particularly pronounced. Although national curricular reforms have emphasized communicative competence and learner-centered instruction, empirical evidence suggests that classroom-based speaking instruction often falls short of its intended outcomes. Large class sizes, limited contact hours, and examination-oriented practices continue to restrict students' opportunities for sustained oral practice and individualized feedback. As a result, many university students demonstrate modest gains in speaking proficiency despite years of formal English instruction, a concern that has been repeatedly highlighted in empirical studies (Nguyen & Le, 2018; Tran, 2020).

The rapid development of artificial intelligence (AI) technologies has opened new possibilities for

Email: Nguyenhonggiang@gmail.com
DOI: 10.64410/MQCT3263

addressing some of these longstanding pedagogical constraints. Recent advances in natural language processing, speech recognition, and conversational agents have enabled the creation of AI-supported learning environments that can simulate interactive communication, provide immediate feedback, and allow learners to engage in repeated speaking practice beyond the classroom. These affordances are particularly relevant for speaking instruction, which requires extensive practice, timely feedback, and opportunities for interaction - conditions that are difficult to ensure consistently in traditional classroom settings.

International research has increasingly explored the pedagogical potential of AI in language education, with a growing body of empirical studies reporting positive effects on learners' oral performance, speaking confidence, and interactional competence. However, existing findings also reveal variability in learning outcomes across different components of speaking ability, such as fluency, accuracy, complexity, and interaction. Moreover, much of the current research has been conducted in non-Vietnamese contexts, and relatively few studies have employed controlled experimental designs within Vietnamese higher education settings. This contextual gap limits the generalizability of existing findings and underscores the need for locally grounded empirical evidence.

In the university that this research is taken place, the integration of digital technologies into English language instruction has been actively encouraged in recent years. Nevertheless, the use of AI specifically to support speaking instruction remains at an early stage, and its effectiveness has not yet been systematically examined through empirical research. Given the institutional emphasis on innovation and quality assurance, there is a clear need for evidence-based evaluation of AI-supported pedagogical interventions, particularly those targeting core language skills such as speaking.

Against this backdrop, the present study investigates the effectiveness of AI-supported speaking instruction in improving English speaking competence among undergraduate students. Drawing on a quasi-experimental design, the study compares the speaking performance of an experimental group receiving AI-assisted speaking practice with that of a control group receiving traditional instruction. Importantly, speaking competence is conceptualized as a multidimensional construct, and learning outcomes are examined across several components, including fluency, accuracy, complexity, and interactional competence.

The study aims to address the following research questions:

1. Is there a significant difference in overall speaking performance between students receiving AI-supported instruction and those receiving traditional instruction?

2. To what extent does AI-supported instruction affect specific components of speaking competence, namely fluency, accuracy, complexity, and interactional competence?

By addressing these questions, the study makes several contributions to the existing literature. First, it provides empirical evidence on the effectiveness of AI-supported speaking instruction in a Vietnamese university context, thereby extending the geographical and contextual scope of current research. Second, by adopting a componential analysis of speaking competence, the study offers a more nuanced understanding of how different aspects of speaking respond to AI-mediated practice. Finally, the findings have practical implications for curriculum design and instructional decision-making in tertiary EFL education, particularly in institutions seeking to integrate AI technologies in a pedagogically principled manner.

2. Literature Review

2.1. Speaking Competence as a Multidimensional Construct

Empirical research in applied linguistics has consistently demonstrated that speaking proficiency is not a unitary ability but a composite construct comprising multiple interrelated dimensions (Bygate, 1987; Ellis & Barkhuizen, 2005; McCrocklin, 2016). Research showed that fluency and accuracy often

develop at different rates and may even compete under certain task conditions (Bibauw, et al., 2019; Housen, Kuiken & Vedder, 2012). The findings indicate that improvements in one dimension, such as fluency, do not automatically entail parallel gains in accuracy or complexity.

It is believed that learners' fluency and interactional effectiveness were more sensitive to increased practice opportunities, whereas complexity exhibited slower and more gradual development (Ellis & Barkhuizen, 2005; Liakin, et al., 2017) and instructional interventions targeting speaking need to be evaluated across multiple dimensions rather than relying on overall speaking scores alone.

Interactional competence has been foregrounded as a distinct and essential component of speaking ability (Skehan, 1998). Learners' ability to manage turn-taking, respond contingently, and sustain interaction significantly predicted communicative success, even when grammatical accuracy was imperfect (Bibauw, François, Desmet, & Chiarcos, 2019; Bygate, 1987). This line of research has important implications for evaluating instructional approaches that emphasize interaction, such as AI-mediated dialogue systems.

A substantial body of empirical research has documented persistent difficulties in teaching and learning speaking in EFL contexts, particularly in higher education. Studies conducted in Asian university settings have consistently reported limited opportunities for oral practice due to large class sizes and time constraints. For example, Nguyen & Le (2018), Tran (2020) found that in many EFL classrooms, individual learners had minimal speaking time, which constrained opportunities for meaningful output and feedback (Nguyen & Le, 2018; Tran, 2020).

Nguyen and Le (2018), using survey and interview data from university students, reported that learners perceived speaking as the most challenging skill and cited lack of practice and fear of making mistakes as major obstacles (Nguyen & Le, 2018). Tran (2020), examining classroom interaction patterns in Vietnamese universities, found that a small number of students tended to dominate speaking activities, while the majority remained passive participants (Tran, 2020).

These findings collectively indicate that traditional classroom-based instruction often fails to provide sufficient conditions for speaking development, thereby motivating the exploration of supplementary instructional approaches, including technology-mediated solutions.

2.2. AI-supported Speaking Practice and Overall Speaking Performance

An increasing number of empirical studies have investigated the impact of AI-supported instruction on overall speaking performance. Bibauw et al. (2019), in a study examining chatbot-mediated interaction among EFL learners, reported significant improvements in learners' oral production after sustained engagement with conversational agents. Their quantitative analysis showed that learners who interacted regularly with chatbots produced longer utterances and demonstrated greater confidence in spoken communication.

Similarly, Li, Kohnke, and Zhang (2023) conducted an experimental study comparing chatbot-supported speaking practice with traditional instruction (Li, Kohnke & Zhang, 2023). Their results indicated that the experimental group significantly outperformed the control group on post-test speaking assessments, with effect sizes suggesting a moderate to strong instructional impact. The authors attributed these gains to increased opportunities for output and reduced speaking anxiety in AI-mediated environments. These studies provide empirical support for the claim that AI-supported speaking instruction can lead to measurable improvements in overall speaking proficiency, a finding that directly corresponds to the significant post-test differences observed in the present study (Lee, Park & Kim, 2022; Li, Kohnke & Zhang, 2023).

2.3. Effects of AI on Fluency and Interactional Competence

Fluency has been identified as one of the speaking components most responsive to AI-mediated practice. Lee, Park & Kim (2022) investigated synchronous computer-mediated communication, found

that learners engaged in technology-mediated interaction produced speech with fewer pauses and greater temporal continuity compared to traditional classroom interaction. More recent studies focusing specifically on chatbots have reported strong gains in interactional competence. Lee et al. (2022) found that learners who practiced speaking with AI conversational agents demonstrated improved turn-taking behavior and greater responsiveness in post-intervention speaking tasks (Lee, Park & Kim, 2022). The authors argued that repeated exposure to dialogic interaction with AI systems helped learners internalize conversational norms and interactional routines. These findings suggests that AI-mediated dialogue is particularly effective in fostering aspects of speaking that rely on sustained interaction and repeated practice (Liakin, Cardoso & Liakina, 2017; Li, Kohnke & Zhang, 2023; Tran, 2020).

2.4. Effects of AI-supported Feedback on Accuracy

Empirical research has also examined the role of AI-generated feedback in improving speaking accuracy. Li et al (2023) believed that immediate, automated feedback had a significant positive effect on learners' grammatical accuracy. Speech recognition technologies, in particular, were found to enhance learners' awareness of pronunciation and form-related errors. Meanwhile, McCrocklin (2016) showed that learners who received automated pronunciation feedback demonstrated greater improvements in segmental accuracy compared to those who relied solely on teacher feedback. These findings support the view that AI tools can complement classroom instruction by providing frequent and timely feedback that is difficult to achieve in large classes (McCrocklin, 2016).

2.5. Mixed Findings on Linguistic Complexity

In contrast to fluency and accuracy, findings related to linguistic complexity have been less consistent. Complexity is believed to be less sensitive to short-term instructional interventions and may require extended exposure or explicit pedagogical scaffolding (Bygate,1987; Housen, Kuiken & Vedder, 2012; Tran, 2020). Studies examining AI-supported instruction have echoed this pattern, only modest gains in syntactic complexity despite significant improvements in fluency and accuracy (Lee, et al., 2022; Li, et al., 2023). While AI-mediated practice increases quantity of output, it does not necessarily prompt learners to experiment with more complex linguistic structures unless guided by specific instructional prompts (Housen, et al., 2012; Lee, et al, 2022). Although prior research provides substantial evidence for the potential of AI-supported speaking instruction, several gaps remain. First, there is a relative scarcity of controlled experimental studies conducted in Vietnamese higher education contexts (Skehan, 1998). Second, many studies report overall speaking gains without systematically analyzing individual speaking components. Third, interactional competence, despite its recognized importance, remains underexplored in AI-related research.

The present study addresses these gaps by employing a quasi-experimental design, focusing on a Vietnamese university context, and conducting a componential analysis of speaking proficiency. By doing so, the study not only corroborates existing findings but also extends them by providing detailed empirical evidence on how different aspects of speaking respond to AI-supported instruction.

3. Methodology

3.1. Research Design

This study adopted a quasi-experimental pre-test–post-test control group design, which is widely employed in applied linguistics research to examine causal relationships between pedagogical interventions and learning outcomes. The choice of this design was motivated by its capacity to capture learning gains over time while controlling for pre-existing differences in learners' proficiency levels.

Two groups of students - an experimental group and a control group -were assessed on their speaking performance before and after the intervention. The pre-test was administered to establish baseline equivalence between the two groups, whereas the post-test was used to determine the extent to which the AI-supported instructional approach contributed to learners' speaking development. By

comparing both within-group gains and between-group differences, the design allowed for a robust evaluation of the instructional impact while minimizing threats to internal validity.

3.2. Participants and sampling

The participants were 70 first-year undergraduate students enrolled in non-English-major programs at a university in Hanoi. The students were between 18 and 20 years of age and had completed secondary-level English education prior to entering university. Based on the institutional placement test administered at the beginning of the academic year, all participants demonstrated English proficiency equivalent to B1 level on the Common European Framework of Reference (CEFR).

Due to institutional constraints, convenience sampling was employed. However, to enhance comparability, participants were randomly assigned to either the experimental group (n = 35) or the control group (n = 35). Random assignment was implemented to reduce selection bias and to ensure that learner-related variables such as prior exposure to English, motivation, and learning strategies were distributed as evenly as possible across the two groups.

3.3. Instruments

3.3.1. Speaking Test

Speaking proficiency was assessed through a semi-structured oral interview, designed to align with the communicative demands of the B1 level and the instructional objectives of the course. The semi-structured format ensured a balance between standardization and authenticity, allowing all participants to respond to comparable prompts while maintaining opportunities for spontaneous language production.

The test comprised three sections. The first section elicited short responses to personal questions to assess basic interactive ability. The second section required participants to describe a visual stimulus or situational prompt, targeting their ability to organize discourse and employ descriptive language. The final section involved a topic-based discussion, which assessed higher-level interactional skills, including turn-taking, elaboration, and opinion expression. All speaking performances were audio-recorded for subsequent rating and analysis.

3.3.2. Rating Rubric and Procedure

An analytic rating rubric was employed to evaluate participants' speaking performance. The rubric consisted of four criteria corresponding to the targeted subcomponents of speaking proficiency: fluency, accuracy, complexity, and interactional competence. Each criterion was rated on a nine-point scale with detailed descriptors to guide rater judgment.

The rubric was adapted from established international speaking assessment frameworks and calibrated to the specific context of Vietnamese tertiary education. The use of an analytic rubric facilitated fine-grained analysis of learners' development across different dimensions of speaking, thereby providing more nuanced insights into the effects of AI-supported instruction.

The study was conducted over an eight-week instructional period, following a structured three-phase procedure.

In the first phase, all participants completed the pre-test under identical conditions. The purpose of this phase was to collect baseline data on students' speaking proficiency and to verify initial equivalence between the experimental and control groups.

The second phase involved the instructional intervention. Both groups followed the same syllabus, used the same course materials, and were taught by the same instructor to control for instructional variability. The experimental group engaged in AI-supported speaking practice outside regular class hours, utilizing conversational chatbots for simulated dialogues and speech recognition tools for pronunciation practice with immediate feedback. Students in this group were encouraged to engage with these tools for approximately 20–30 minutes per week. In contrast, the control group received

conventional speaking instruction, relying primarily on in-class pair and group activities without access to AI-based tools.

The final phase consisted of the post-test, which mirrored the structure and difficulty level of the pre-test. This ensured that any observed differences in performance could be attributed to instructional effects rather than test-related factors.

Two experienced university-level English instructors independently rated all speaking samples. Prior to formal rating, the raters participated in a calibration session to discuss the rubric descriptors and align their scoring standards. Inter-rater reliability was calculated using correlation analysis, indicating an acceptable level of agreement between the two raters. Discrepancies were resolved through discussion to reach a consensus score.

3.4. Data Analysis

All quantitative data were analyzed using SPSS version 26.0. Descriptive statistics, including means and standard deviations, were first computed to summarize participants' speaking performance. Inferential analyses included independent-samples t-tests to examine between-group differences and paired-samples t-tests to assess within-group gains from pre-test to post-test. Statistical significance was set at $p < .05$. Additionally, Cohen's d was calculated to estimate effect sizes, providing an index of the practical significance of the AI-supported intervention.

4. Results

4.1. Comparison of Pre-test Speaking Performance Between the Experimental and Control Groups

Table 1. Comparison of pre-test of experimental and control group

Group	N	Mean	SD	t	p
Experimental	35	5.62	0.68		
Control	35	5.59	0.71	0.18	.857

An independent-samples t-test was conducted to examine whether there were any statistically significant differences between the experimental group and the control group in overall speaking performance prior to the instructional intervention. As shown in Table 4.,1, the mean pre-test speaking score of the experimental group ($M = 5.62$, $SD = 0.68$) was highly comparable to that of the control group ($M = 5.59$, $SD = 0.71$)

The statistical analysis revealed no significant difference between the two groups at the pre-test stage ($t = 0.18$, $p = .857$). The very small t-value and the non-significant p-value indicate that any observed difference in mean scores was negligible and likely due to random variation rather than systematic differences between groups. In addition, the similar standard deviation values suggest that the degree of score dispersion was comparable across the two groups.

These results confirm that the experimental and control groups were statistically equivalent in terms of overall speaking proficiency prior to the intervention, thereby establishing a valid baseline for subsequent comparisons of post-intervention outcomes.

4.2. Comparison of Post-Test Speaking Performance Between the Experimental and Control Groups

Table 2. Comparison of post-test of experimental and control group

Group	N	Mean	SD	t	p	Cohen's d
Experimental	35	6.98	0.65			
Control	35	6.21	0.69	4.71	< .001	1.12

To assess the impact of AI-supported speaking instruction, an independent-samples t-test was performed to compare the post-test speaking scores of the experimental and control groups. The descriptive statistics and inferential results are presented in Table 4.2.

The results show that the experimental group achieved a substantially higher mean post-test speaking score (M = 6.98, SD = 0.65) than the control group (M = 6.21, SD = 0.69). The difference between the two groups was statistically significant (t = 4.71, p < .001), indicating that the likelihood of this difference occurring by chance is extremely low.

In addition to statistical significance, the effect size was calculated using Cohen's d to estimate the magnitude of the instructional effect. The obtained effect size (d = 1.12) can be interpreted as large according to conventional benchmarks, suggesting that the AI-supported instructional approach exerted a strong practical impact on students' overall speaking performance.

4.3. Within-group Development in the Experimental Group

Table 3. Results of paired t-test of the experimental group

Components	Pre-test mean	Post-test mean	t	p
Fluency	5.60	7.10	7.84	< .001
Accuracy	5.75	6.85	6.12	< .001
Complexity	5.52	6.10	2.34	< .005
Interaction ability	5.61	7.25	8.21	< .001

To further examine how the experimental group's speaking performance developed over time, paired-samples t-tests were conducted to compare pre-test and post-test scores across individual speaking components. The results are summarized in Table 4.3.

The analysis revealed statistically significant improvements across all four speaking components. For fluency, the mean score increased from 5.60 at the pre-test to 7.10 at the post-test, with the difference reaching a high level of statistical significance (t = 7.84, p < .001). This result indicates a substantial improvement in students' ability to speak more smoothly and with fewer hesitations.

Accuracy also showed a significant increase, with mean scores rising from 5.75 to 6.85 (t = 6.12, p < .001). This suggests that students demonstrated greater control over grammatical structures and lexical choices following the intervention. Complexity exhibited a more moderate but still statistically significant improvement, increasing from a mean of 5.52 to 6.10 (t = 2.34, p <.005). Although the magnitude of change in complexity was smaller than that observed for fluency and accuracy, the result nonetheless indicates measurable development in the use of more varied linguistic structures.

The most pronounced improvement was observed in interactional competence, where the mean score increased from 5.61 to 7.25. This difference was highly significant (t = 8.21, p < .001), reflecting substantial gains in students' ability to initiate, sustain, and appropriately respond during spoken interaction.

Taken together, these results demonstrate that the experimental group experienced consistent and statistically significant improvement across all measured components of speaking proficiency during the intervention period.

It can be seen that the results indicate that the experimental and control groups began the study with statistically equivalent levels of speaking proficiency. Following the instructional intervention, the experimental group significantly outperformed the control group in overall speaking performance, with a large effect size observed at the post-test. Within-group analyses further revealed that the experimental group made significant gains across all major components of speaking, particularly in fluency and interactional competence.

5. Discussion

5.1. Influence of AI-supported Instruction in EFL Students' Speaking Performance at the Tertiary Level

The findings of the present study provide clear empirical evidence supporting the effectiveness of AI-supported instruction in enhancing EFL students' speaking performance at the tertiary level. The statistically significant difference in post-test speaking scores between the experimental and

control groups, accompanied by a large effect size (Cohen's d = 1.12), indicates that the integration of AI tools into speaking instruction led to substantial learning gains beyond those achieved through conventional teaching methods.

This finding is consistent with previous international studies that have reported positive effects of AI-mediated interaction on oral language development. For example, research by Li, Kohnke, and Zhang (2023) and Bibauw et al. (2019) demonstrated that AI-based conversational agents provide learners with increased opportunities for output, immediate feedback, and low-anxiety practice environments, all of which contribute to improved speaking performance. The current study extends these findings to the Vietnamese tertiary context, where empirical evidence on AI-supported speaking instruction remains limited. Though Nguyen & Le (2018), Tran (2020) mentioned the task of teaching speaking skill in tertiary context with the help of technology, it is hard to deny that there should be more scientific results to position the role of AI in learning a foreign language.

5.2. Baseline Equivalence and Validity of the Observed Gains

The absence of statistically significant differences between the experimental and control groups at the pre-test stage strengthens the internal validity of the study. The near-identical mean scores and comparable standard deviations suggest that both groups started with equivalent speaking proficiency, thereby reducing the likelihood that post-test differences resulted from pre-existing disparities. This methodological rigor aligns with best practices in experimental research in applied linguistics (Tran, 2020). By establishing baseline equivalence, the study ensures that the observed gains in the experimental group can be more confidently attributed to the AI-supported instructional intervention rather than extraneous variables such as learner proficiency, motivation, or prior exposure to English.

5.3. Differential Development Across Speaking Subcomponents

One of the most notable contributions of this study lies in its fine-grained analysis of speaking subcomponents. The within-group analysis of the experimental group revealed statistically significant gains across all four dimensions of speaking proficiency, although the magnitude of improvement varied among components.

The strongest gains were observed in fluency and interactional competence. This pattern aligns with previous studies suggesting that AI-mediated interaction is particularly effective in fostering aspects of speaking that benefit from frequent practice and sustained interaction. Technology-mediated environments encourage learners to produce longer stretches of speech and engage more actively in conversational exchanges (McCrocklin, 2016; Nguyen & Le, 2018). The substantial improvement in interactional competence may be attributed to the dialogic nature of chatbot-based practice, which requires learners to initiate turns, respond appropriately, and maintain coherence in extended interactions. This finding echoes the work of Bibauw et al. (2019), who emphasized the role of conversational agents in developing pragmatic and interactional skills that are often underemphasized in traditional classroom instruction.

5.4. Gains in Accuracy and Complexity: Convergence and Divergence with Previous Research

The significant improvement in accuracy observed in the experimental group is consistent with studies highlighting the benefits of immediate feedback provided by AI tools. Speech recognition technologies and AI-generated corrective feedback may have heightened learners' noticing of linguistic errors, thereby facilitating form-focused learning alongside communicative practice. This finding supports that automated feedback can contribute to measurable gains in grammatical accuracy (Lee, Park & Kim, 2022). In contrast, while complexity also improved significantly, the magnitude of change was more modest compared to fluency and interactional competence. This result is in line with previous research suggesting that syntactic and lexical complexity tends to develop more gradually and may require longer instructional periods or more explicit pedagogical support. Complexity gains are often less pronounced in short-term interventions.

The present findings therefore both converge with and extend existing literature by demonstrating that AI-supported instruction can positively influence complexity, albeit to a lesser extent than other speaking components within an eight-week timeframe.

5.5. Comparison of Mean Gains and Instructional Implications

The comparison of mean gain scores between the experimental and control groups provides further support for the effectiveness of AI-supported instruction. The substantial gain observed in the experimental group, contrasted with the negligible improvement in the control group, suggests that traditional classroom-based speaking activities alone may be insufficient to promote significant speaking development within a limited instructional period.

This result aligns with Vietnamese studies that have highlighted constraints in conventional speaking instruction, including limited class time and unequal participation opportunities (Nguyen & Le, 2018; Tran, 2020). By offering additional practice opportunities outside the classroom, AI tools appear to address these constraints and supplement face-to-face instruction in meaningful ways.

Despite these contributions, the study has several limitations. The relatively short duration of the intervention may have constrained the extent of development observed in certain aspects of speaking, particularly linguistic complexity. In addition, the sample size and institutional context limit the generalizability of the findings. Future research could adopt longitudinal designs to examine the long-term effects of AI-supported speaking instruction and incorporate qualitative methods to explore learners' experiences and interactional processes in greater depth.

6. Conclusion

This study investigated the effectiveness of integrating artificial intelligence into speaking instruction in a Vietnamese tertiary EFL context. Employing a quasi-experimental pre-test–post-test control group design, the research examined whether AI-supported practice could enhance students' speaking performance beyond what is achieved through conventional instruction. The findings provide compelling evidence that AI-supported speaking instruction leads to significant and meaningful improvements in learners' oral proficiency.

The present study contributes to the growing body of research on AI in language education in several important ways. First, it provides robust quantitative evidence from a controlled experimental design in a Vietnamese university context, which remains underrepresented in the international literature. Second, by examining multiple dimensions of speaking proficiency, the study offers a more nuanced understanding of how AI-supported instruction affects different aspects of oral language development. Finally, the large effect size observed in this study suggests /that AI integration, when systematically implemented, can yield pedagogically meaningful outcomes rather than marginal gains

The results demonstrate that while the experimental and control groups began the study with comparable levels of speaking proficiency, only the experimental group exhibited substantial improvement over the instructional period. Students who engaged in AI-supported speaking practice significantly outperformed their peers in overall speaking performance, with a large effect size observed at the post-test. This outcome confirms that the integration of AI tools is not merely a supplementary feature but can function as a powerful pedagogical intervention when systematically embedded in speaking instruction.

A key contribution of this study lies in its componential analysis of speaking proficiency. The experimental group showed significant gains across all major dimensions of speaking, including fluency, accuracy, complexity, and interactional competence. Particularly strong improvements were observed in fluency and interactional competence, suggesting that AI-mediated environments are especially conducive to developing aspects of speaking that require frequent practice, sustained interaction, and reduced anxiety. These findings extend existing research by providing empirical support from a Vietnamese university context, where studies on AI-supported speaking instruction remain limited.

From a pedagogical perspective, the findings underscore the potential of AI tools to address persistent challenges in speaking instruction, such as limited classroom time and unequal participation opportunities. By offering learners additional, flexible, and low-pressure speaking practice outside the classroom, AI-supported instruction can complement traditional teaching and promote more equitable access to oral practice. For instructors and curriculum designers, the results suggest that AI tools should be integrated in a principled and pedagogically informed manner, rather than used as isolated or optional add-ons.

This study provides robust empirical evidence that AI-supported instruction can significantly enhance EFL learners' speaking performance at the tertiary level. By demonstrating both statistically significant and pedagogically meaningful gains, the study contributes to the growing body of research on artificial intelligence in language education and offers practical implications for improving speaking instruction in higher education.

The research still has its own weaknesses. The time for the intervention is quite short (eight weeks), which may not fully reflect the actual effects that students may receive. In the future, longer time should be assigned to the course. The number of students involved is also small, therefore this result cannot represent the teaching method or pedagogical advantages of AI in general, rather it can be seen as a case study.

Conflicts of interest: No conflict of interest.

References

Bibauw, S., François, T., Desmet, P., & Chiarcos, C. (2019). *Computer assisted language learning and natural language processing: Needs, limits and prospects.* Proceedings of the 8th Workshop on NLP for Computer Assisted Language Learning, 1–13.

Bygate, M. (1987). *Speaking.* Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford University Press.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Complexity, accuracy and fluency: Definitions, measurement and research*. Studies in Second Language Acquisition, 34(1), 1–20. https://doi.org/10.1017/S0272263111000485

Lee, S., Park, M., & Kim, H. (2022). *Effects of AI-based chatbots on EFL learners' speaking performance and anxiety. Computer Assisted Language Learning.* Advance online publication. https://doi.org/10.1080/09588221.2022.2040536

Li, R., Kohnke, L., & Zhang, D. (2023). *Artificial intelligence in language learning: Chatbots and oral skill development.* System, 113, 102979. https://doi.org/10.1016/j.system.2022.102979

Liakin, D., Cardoso, W., & Liakina, N. (2017). *Learning L2 pronunciation with a mobile speech recognizer: French /y/.* CALICO Journal, 34(1), 1–25. https://doi.org/10.1558/cj.29562

McCrocklin, S. (2016). *Pronunciation learner autonomy: The potential of automatic speech recognition.* System, 57, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Nguyen, T. M. H., & Le, T. T. H. (2018). *Students' perceptions of difficulties in learning speaking skills at Vietnamese universities.* VNU Journal of Foreign Studies, 34(2), 45–58. https://js.vnu.edu.vn/FS/article/view/4185

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Tran, T. Q. (2020). *Classroom interaction patterns in EFL speaking lessons at Vietnamese universities*. Journal of Science, Ho Chi Minh City University of Education, 17(4), 45–56. https://journal.hcmue.edu.vn/index.php/hcmuejos/article/view/1846.