

Improving English Communication Skills for Non-English Major EFL Students Using a Self-Developed Website - *Speakee*

Nguyen Le Nguyen¹, Pham Thi Thao Nguyen¹, Uyen Nu Thuy Nguyen^{1*}

¹ Faculty of English, University of Foreign Language Studies, The University of Da Nang, Vietnam

* Corresponding author's email: ngntuyen@ufl.udn.vn

 <https://orcid.org/0000-0001-5798-1784>

 <https://doi.org/10.54855/ijte.26644>

® Copyright (c) 2026 Nguyen Le Nguyen, Pham Thi Thao Nguyen, Uyen Nu Thuy Nguyen

Received: 05/07/2026

Revision: 19/08/2026

Accepted: 09/09/2026

Online: 14/10/2026

ABSTRACT

Keywords: self-developed speaking website; *Speakee*; English communication skills; non-English major EFL students

The study examined the effects of a self-developed website – *Speakee* on developing English communication skills among non-English major students at the University of Da Nang. It also examined how students perceived the use of the website as a tool for improving their English communication skills. *Speakee* is an AI-powered role-play platform developed to provide contextualized speaking practice and delayed feedback, which was designed based on Task-Based Language Teaching and Situated Learning Theory. The study adopted a mixed-methods quasi-experimental design. Data were gathered through English speaking tests during pre- and post-intervention and semi-structured interviews. The results indicated that students using *Speakee* demonstrated significant improvement in their overall English-speaking performance. Interview findings further suggested that students perceived several benefits from using the platform, particularly in pronunciation practice, vocabulary development, speaking confidence, and opportunities for additional speaking practice. Nevertheless, several technical limitations related to speech recognition accuracy, AI interaction quality, and long-term engagement were also reported.

Introduction

English communication proficiency is vital for global integration and employment. Although the Vietnamese government has prioritized establishing English language learning at the tertiary level, practical implementation remains a significant challenge. Despite greater opportunities to learn English, many still struggle to reach the level of proficiency expected in real-life communication (Ngo & Tran, 2023). This gap is particularly clear in English communication, where many learners struggle to apply grammatical knowledge to real-life communication. Indeed, English speaking is often considered particularly difficult, often causing high levels of anxiety among EFL learners (Kessler, 2010). In Asian countries, outdated EFL classroom approaches have not been sufficiently updated or reformed, leaving many learners with limited opportunities to practice speaking and engage in interactive English

activities (Baek & Lee, 2018; Chen & Hwang, 2019). In Vietnam, this issue worsens due to a lack of practice environments outside the classroom and tailored materials for students not majoring in English (Trinh & Pham, 2021). These situations could result in graduates' difficulties in English communication in future workplace settings.

Recent studies have shown that Artificial Intelligence (AI) systems such as chatbots and virtual assistants can improve communication skills by providing interactive and immediate feedback, enhancing fluency, pronunciation, confidence, and opportunities for speaking practice (Kohnke et al., 2023), and fostering learner engagement and lowering psychological barriers (Vasu et al., 2025). Nevertheless, custom-built platforms designed specifically for non-English major EFL students remain scarce. Few platforms provide workplace-related communication scenarios tailored to the disciplinary needs of learners, as most existing applications provide general content that might not resemble authentic communication contexts.

To address this gap, the researchers developed an AI-powered role-play website, named *Speakee*, that integrates contextualized role-play scenarios with structured feedback. The study aimed to fulfill two objectives. First, it examined the effects of using the *Speakee* website on non-English major EFL students' overall English-speaking performance. Second, it explored the benefits and drawbacks that students faced when using *Speakee* to improve their English communication skills. The study's findings would provide implications for AI-assisted and learner-centered speaking practice tools. Unlike applications that mainly provide predetermined topics, *Speakee* lets learners create customized role-play scenarios based on their communication needs and provides structured speaking practice and feedback.

Research Questions

The study sought to answer two research questions:

1. To what extent does the self-developed website *Speakee* enhance the English communication skills of non-English major EFL students at the University of Da Nang?
2. What benefits and drawbacks did non-English major EFL students perceive when using *Speakee* for their communication skills development?

Literature review

AI Use and English Communication Skills

English communication skills refer to the capacity to use English to appropriately express thoughts, feelings, and information (Rao, 2019). Instead of concentrating only on knowledge of vocabulary and grammar, this concept focuses on language use in authentic communicative situations. This perspective aligns with Hymes's (1972) theory of communicative competence, which holds that communication involves grammatical correctness and adapting language to social settings. This study adopts Rao's (2019) definition because it emphasizes language usage in context rather than memorization.

Previously, AI-related research often examined AI as tutoring systems guiding learners, but the current focus has shifted towards generative AI technologies, which can produce personalized and interactive learning materials adapting to learner needs (Kohnke et al., 2023). These technologies can support English pedagogy, including vocabulary growth, grammar, pronunciation, speaking fluency and coherence, and overall speaking performance (Chen et al., 2025; Du & Daniel, 2024; Fathi et al., 2024; Shan et al., 2024). Furthermore, AI platforms

promote learner autonomy by providing a self-paced environment with diagnostic feedback and scaffolding that helps learners identify and improve pronunciation and intonation problems (Chan & Hu, 2023). These AI tools also lower speaking anxiety, boost students' willingness to communicate, and increase learner motivation and autonomy (Fathi et al., 2024; Tai & Chen, 2023). In the Vietnamese higher-education context, Ho (2024) also examined university students' attitudes, habits, and perceptions regarding ChatGPT use in English language learning, further illustrating the growing role of AI-assisted tools in tertiary EFL learning. Recent research on automatic speech recognition (ASR) also indicates that ASR-supported practice can benefit pronunciation development, while recognition and feedback accuracy remain important considerations in EFL speaking contexts (Ngo et al., 2024; Sun, 2023).

Theoretical Framework

This study was grounded in three relevant theoretical perspectives: Task-Based Language Teaching (TBLT), as conceptualized by Ellis (2003, 2009); Situated Learning Theory (SLT), developed by Lave and Wenger (1991); and the Communicative Competence (CC) framework, proposed by Canale and Swain (1980) and subsequently refined by Canale (1983). Firstly, TBLT focuses on utilizing language to accomplish meaningful communication goals. Instead of concentrating mainly on isolated grammatical structures, it encourages learners to apply language within contextualized situations to complete specific tasks. According to Ellis (2003), a task requires learners to use language purposefully in order to complete a specific communicative goal. The main features of TBLT include an emphasis on meaning rather than individual grammatical forms, a focus on tasks that reflect real-world situations, and the use of task completion as the main indicator of success (Ellis, 2009). In practice, TBLT follows the three-stage framework introduced by Willis (1996): pre-task, task performance and reporting, and language focus, which examines and improves the linguistic elements used throughout the task. Recent evidence from Vietnamese non-English-major university students further indicates that TBLT can positively support speaking fluency through task-based practice (Pham, 2025).

Secondly, Situated Learning Theory (SLT) by Lave and Wenger (1991) links learning to the activities and environments in which it takes place, rather than to processes occurring solely in the mind. A central concept of this theory is Legitimate Peripheral Participation, in which learners begin as observers and gradually become active participants within a community (Lave & Wenger, 1991). From this perspective, learning is closely connected to participation in meaningful activities and authentic social contexts. In the present study, SLT informs the development of the website's task content by ensuring that speaking practice activities are based on authentic and everyday situations. These activities are created to represent a range of contexts, including workplace environments and daily social communication. Through these simulated practice opportunities, the website enables non-English major EFL students to learn while building confidence in situations relevant to their future professions and everyday experiences. This enables students to develop communication skills in realistic contexts rather than relying solely on their knowledge of grammatical rules.

Finally, this study uses the Communicative Competence framework established by Canale and Swain (1980) and later refined by Canale (1983). This framework offers a foundation to define communication skills as a measurable construct instead of an abstract concept. The model categorizes language proficiency into four components: grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. Within the present study, the three theoretical perspectives serve complementary roles. TBLT informs the design of goal-oriented speaking tasks that require learners to develop and extend ideas, while SLT supports the use of authentic scenarios that elicit context-relevant vocabulary and

spontaneous spoken responses. The Communicative Competence framework provides a linguistic basis for interpreting grammar and vocabulary use, while the speaking assessment rubric operationalizes pronunciation and idea development. Together, these perspectives link *Speakee*'s design to the dimensions used to evaluate learners' speaking performance.

Review of Previous Relevant Studies

Several studies have examined how AI-supported or self-developed digital platforms could influence English learners' speaking performance in international contexts. Recently, Chen et al. (2025) incorporated a Generative AI agent into role-play activities at a Chinese university. The findings indicated that AI-mediated role-play activities could be beneficial in facilitating learners' speaking performance. Students receiving AI-supported intervention demonstrated higher intrinsic motivation and self-confidence than their counterparts in the other group. No significant difference emerged between learners who practiced role-play with chatbot support and those who participated in conventional role-play activities.

Similarly, Mingyan et al. (2025) examined the influence of *Liulishuo*, an AI-assisted mobile application that provides automated feedback, progress monitoring, and personalized guidance, on Chinese university EFL students' speaking development. The findings showed that learners who used the application improved their speaking performance more than those in the control group, with notable gains in pronunciation and fluency. However, in terms of vocabulary and grammar, no notable improvement was recorded. Beyond linguistic gains, Park et al. (2026) developed *AI Twin*, a personalized conversational system that provided implicit feedback by reformulating learners' utterances into more fluent English. The study showed higher levels of emotional engagement and motivation compared with conventional explicit correction approaches, highlighting the merits of personalized AI support in speaking practice.

In Vietnamese contexts, some studies have examined the utilization of AI in developing English speaking skills in general or aspects of speaking skills in particular. Notably, Nguyen and Nguyen (2023) explored the benefits and challenges that 10 interviewed non-English major students faced when using *Elsa Speak* for developing pronunciation through semi-structured interviews. Students said they appreciated several aspects of the app, including varied exercises, meaningful feedback, detailed error correction, pronunciation guidance, and a user-friendly interface. Nevertheless, they were uncomfortable with the app's repetitive word repetition. Later, Duong and Suppasetseree (2024) explored the effectiveness of *Andy*, an English-speaking chatbot, to improve the English speaking skills of 30 non-English major undergraduate students at Can Tho University, using pre- and post-speaking tests and semi-structured interviews. The results revealed noteworthy improvements in their English speaking skills, especially vocabulary, grammar use, and fluency, and students generally reported positive perceptions of their experience using the *Andy* chatbot.

Although previous studies have demonstrated the potential of AI-supported tools for English speaking development, several gaps remain. First, many studies conducted in the Vietnamese context have examined commercially available applications such as *ELSA Speak* or existing chatbots that mainly provide general or predetermined speaking content. Second, relatively limited attention has been given to locally developed platforms designed specifically for non-English major students and their future communication needs. In addition, few studies have examined platforms that allow learners to actively generate their own role-play scenarios rather than only selecting from pre-designed topics. To address these gaps, *Speakee* was developed to provide voice-based interaction, customized scenarios, structured feedback, progress tracking, and gamified features within a single platform. In particular, its role-play activities can be adjusted according to learners' academic and workplace communication needs.

Methods

Design of the Study

A quasi-experimental mixed-methods design was used to investigate Speakee's effects on the English communication skills of non-English major EFL students and their perceptions of its benefits and drawbacks. For Research Question 1, we analyzed pre-test and post-test scores from English speaking tests to assess Speakee's effects on students' English oral proficiency. For Research Question 2, interview data were analyzed to investigate students' experiences with the platform, including their perceived benefits and limitations.

Participants

This study adopted a purposive sampling strategy, selecting participants based on a set of predetermined criteria. All participants were non-English major EFL students enrolled at the University of Da Nang during the 2025-2026 academic year. Their English proficiency was at the A2 level, as indicated by their enrollment in general English courses at their respective institutions. The participants were enrolled in engineering, information technology, and related disciplines, where general English courses are foundational. The study was conducted with 60 students, divided into a control group ($n = 30$) and an experimental group ($n = 30$). The students shared comparable learning backgrounds. While the control group followed their regular curriculum without accessing the platform, the experimental group was instructed to utilize the Speakee website throughout the intervention period. Subsequently, 15 students from the experimental group were selected through convenience sampling for the semi-structured interviews based on their availability and willingness to share their experiences of using Speakee.

Data Collection

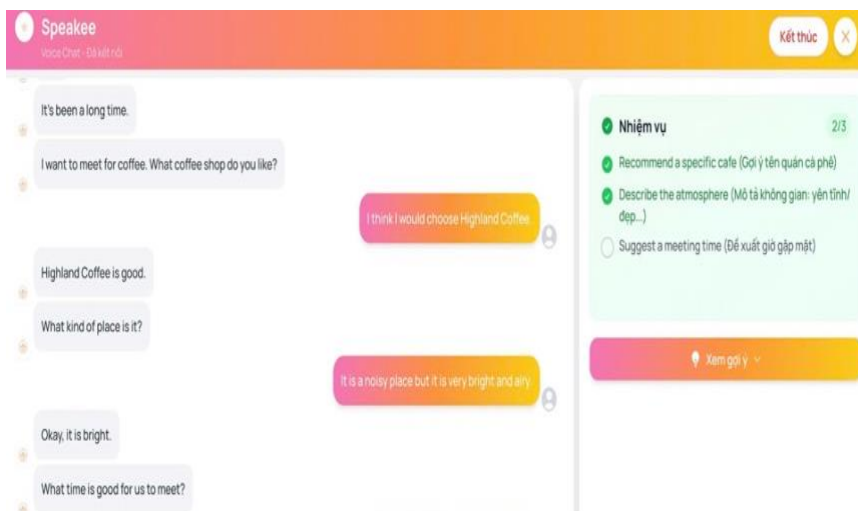
The data collection process consisted of three stages: design, intervention, and post-intervention. In the first stage, the researchers developed a self-developed website called *Speakee*. The platform functioned as both the learning environment and the main data collection tool. *Speakee* is a voice-based learning platform that helps learners develop their English communication skills by providing interactive speaking practice. Students were able to access the *Speakee* platform through the following website: <https://Speakee.dutai.site/auth>.

The platform was developed based on the principles of Task-Based Language Teaching (TBLT) and Situated Learning Theory. These pedagogical foundations emphasize learning through authentic tasks and context-specific interactions. Accordingly, *Speakee* functions as an intelligent conversational agent that uses Generative AI to create dynamic scenarios resembling real-life communication situations. This allows learners to engage in personalized speaking tasks that reflect their own communicative needs and contexts.

The platform consists of four main features, each developed based on established educational principles. The first feature is direct voice-based interaction, the platform's central component, and it reflects TBLT principles. By integrating Speech-to-Text technology and Large Language Models (LLMs), learners can participate in natural, real-time conversations. Consistent with Ellis (2009), this feature emphasizes language use for communication purposes and encourages learners to focus on expressing meaning through spontaneous exchanges.

Figure 1.

Active chat interface



The second feature is the customized scenario generator, which is informed by SLT and is intended to increase the authenticity of language practice. By utilizing AI, the platform generates scenarios that are relevant to learners' individual needs and contexts. According to Lave and Wenger (1991) and Lang (2024), learning can be strengthened through situations resembling real-life experiences.

Figure 2.

"My Scenarios" interface

The screenshot shows a form titled 'Tạo Kịch Bản Tùy Chỉnh' (Custom Scenario Generator). The form includes the following fields and controls:

- Tiêu đề kịch bản ***: A dropdown menu with the placeholder text 'Chọn hoặc nhập tiêu đề kịch bản...'. Below it, a note says 'Chọn một gợi ý hoặc tự nhập tiêu đề của bạn (5-100 ký tự)'.
- Vai trò của AI ***: A text input field with the placeholder text 'VD: Tư vấn viên hướng nghiệp chuyên nghiệp'. Below it, a note says '3-50 ký tự'.
- Vai trò của bạn ***: A text input field with the placeholder text 'VD: Sinh viên mới tốt nghiệp'. Below it, a note says '3-50 ký tự'.
- Thông tin bổ sung (bối cảnh)**: A large text area with the placeholder text 'Thêm bất kỳ bối cảnh hoặc chi tiết bổ sung nào...'. Below it, a note says 'Tối đa 500 ký tự' and '0/500'.
- Mục tiêu cần đạt**: A text input field with the placeholder text 'Mục tiêu 1'. Below it, a note says '2/4 mục'.

At the bottom of the form, there are two buttons: 'Hủy bỏ' (Cancel) and 'Tạo kịch bản' (Generate Scenario).

The third feature is delayed feedback, which corresponds to the post-task stage of the TBLT framework. This stage contributes to language development by allowing learners to review their

performance (Long, 1996). Following each interaction, the platform provides prompt feedback using criteria comparable to those applied in Cambridge assessments, enabling learners to recognize areas that require improvement.

Figure 3.

Overall feedback summary

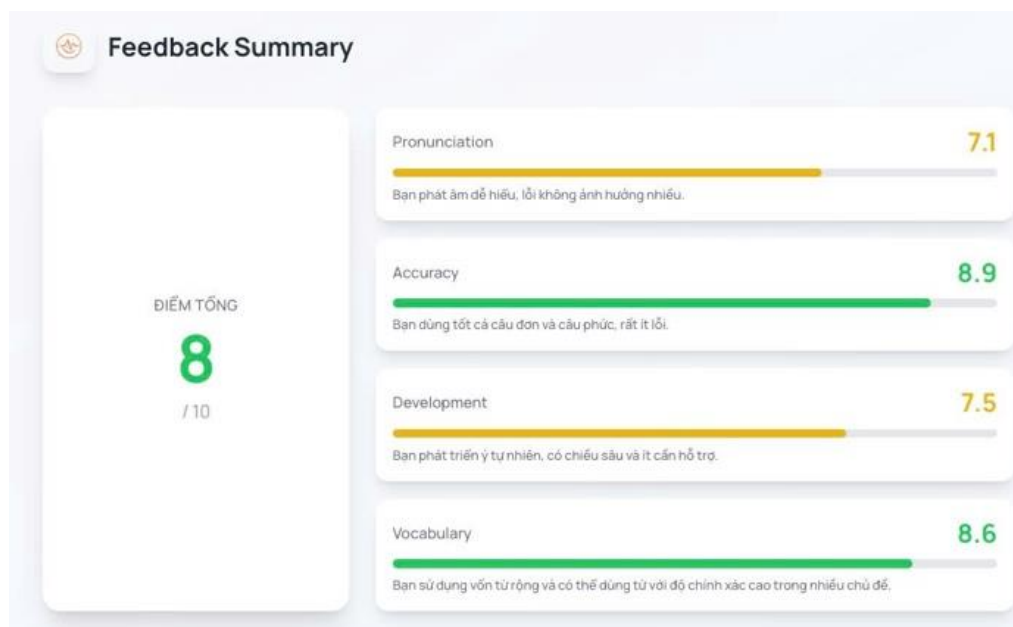
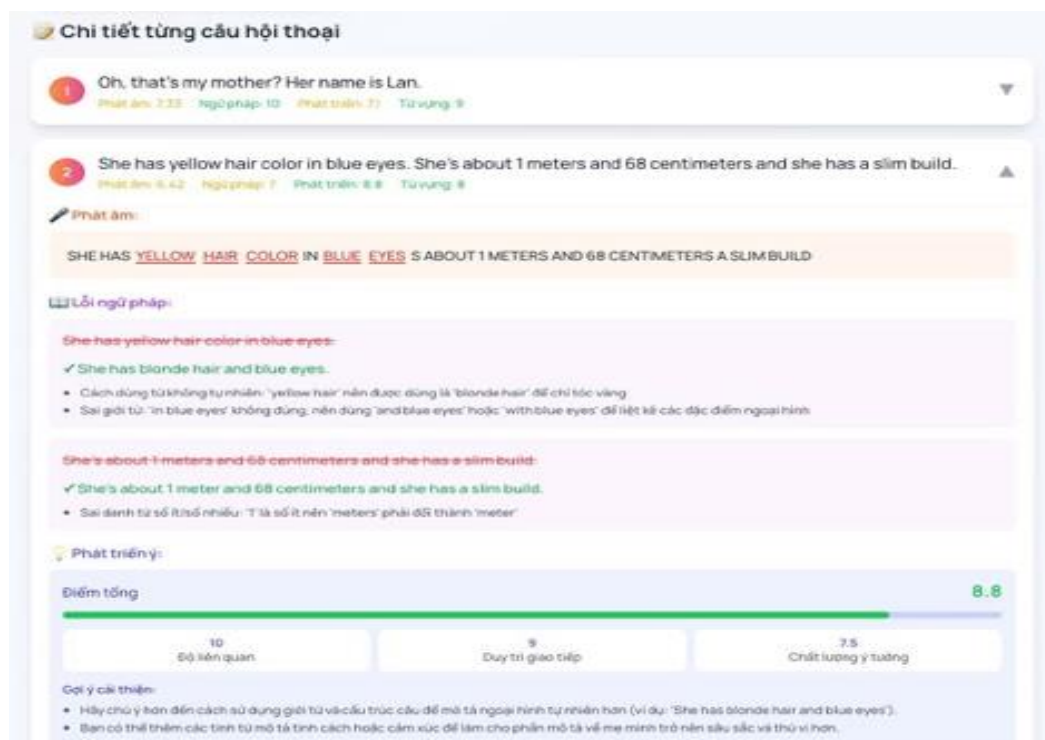


Figure 4.

Detailed feedback



The final feature combines progress tracking with gamification elements to maintain learner participation. *Speakee* monitors users' performance over time and enables them to observe their learning progress. To increase motivation, the platform includes a dragon raising feature in which a virtual character develops as learners complete activities and improve their performance. Deterding et al. (2011) suggested that integrating game elements into non-game settings can support user motivation.

Figure 5.

Learning progress statistics

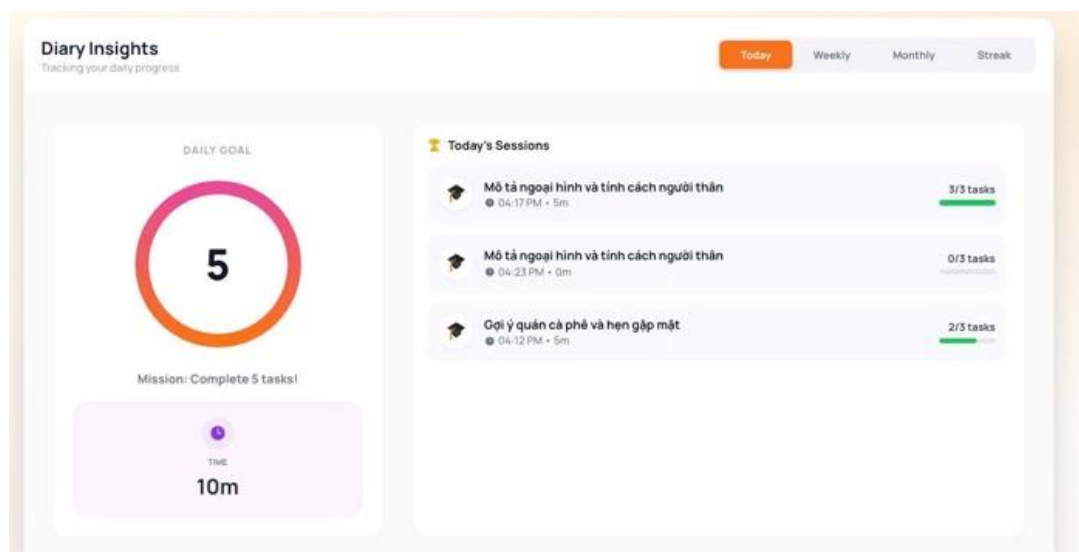
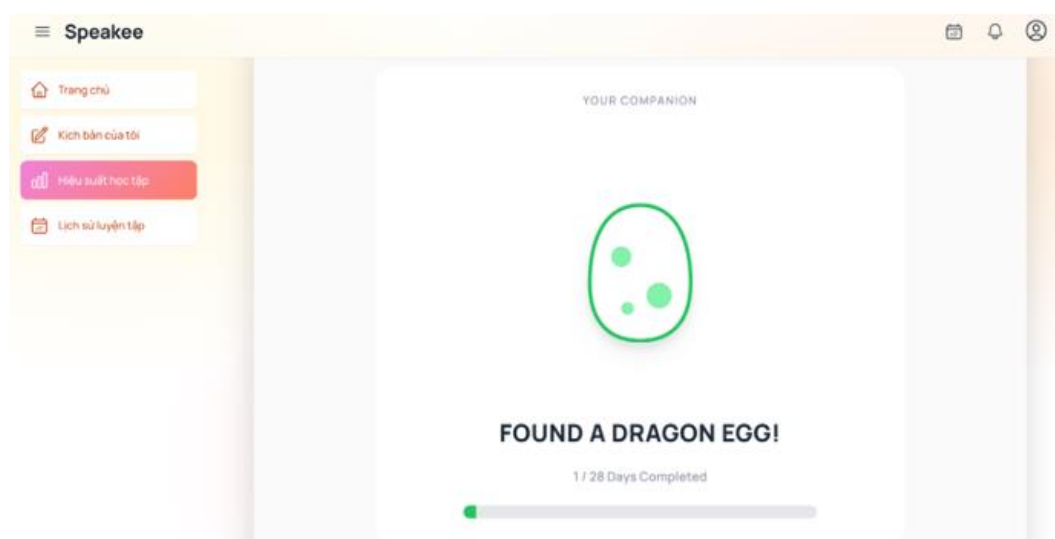


Figure 6.

Gamified dragon feature



It includes a login system to record students' frequency and duration of use, provides simulated academic and workplace communication scenarios, and allows learners to create their own speaking situations for practice.

Before the intervention, all participants completed an oral English speaking test to evaluate their initial English communication proficiency. The test was developed based on the Cambridge English Test for A2 learners, and their performances were assessed using the Cambridge Speaking Assessment Rubric. The six-week intervention began with orientation sessions, during which students received instructions and demonstrations on how to use *Speakee*. Afterward, they practiced independently outside the classroom. No fixed number of sessions or weekly practice time was imposed on all participants; instead, the system recorded individual interaction frequency and duration, which varied across learners. During this period, the system automatically collected usage data, including interaction frequency, duration, and platform-generated performance scores.

After the intervention, students took a post-test with the identical format and assessment criteria. All speaking responses were audio-recorded and independently scored by two raters using the Cambridge Speaking Assessment Rubric to ensure scoring reliability. In order to further explain the quantitative results, 15 participants selected through convenience sampling were invited to join online semi-structured interviews in Vietnamese. The interviews explored the perceived benefits and drawbacks of *Speakee*, including its influence on speaking development, practice opportunities, motivation, communication anxiety, role-play scenarios, feedback, gamification, technical performance, and interaction quality. The interview data were analyzed to provide deeper insights into students' perceptions and further explain the quantitative results.

Data Analysis

The Cambridge Speaking Assessment Rubric assessed oral communication skills using four criteria: pronunciation, vocabulary, grammar, and idea development, with idea development referring to the ability to extend and elaborate responses with relevant content. To ensure scoring reliability, two trained raters with C1 English proficiency independently assessed all recordings, and any scoring discrepancies were discussed and resolved before final scores were determined. Inter-rater reliability was examined using a two-way random-effects Intraclass Correlation Coefficient with an absolute-agreement definition [ICC(2,1)], showing very high agreement for both the pre-test (ICC = .997) and post-test (ICC = .968). The quantitative data were analyzed using SPSS version 26 and Excel 2025. Descriptive statistics were used to summarize students' performance, while an Independent Samples t-test examined pre-intervention equivalence between the control and experimental groups. Because baseline equivalence was not established, ANCOVA was used to compare post-test scores while controlling for pre-test performance. A paired-samples t-test was then used to measure changes within each group after the six-week intervention. Interview data were analyzed using thematic analysis following Braun and Clarke (2006), including familiarization with the transcripts, initial coding, theme development, review, definition and naming, and reporting. Related codes were subsequently grouped into overarching themes.

Results

Research question 1: To what extent does the self-developed website *Speakee* enhance the English communication skills of non-English major EFL students at the University of Da Nang?

An independent-samples t-test was conducted to compare the pre-test scores of the control and experimental groups. The results indicated a significant difference between the two groups, with the control group achieving a higher mean score ($M = 7.633$, $SD = 1.141$) than the experimental group ($M = 6.700$, $SD = 1.132$), $t(58) = 3.180$, $p = .002$. Although the assumption of equal variances was satisfied ($F = 0.026$, $p = .873$), baseline equivalence was not established.

Table 1

Independent-samples t-test results for pre-intervention scores

		Levene's Test (F)	Levene's Test (Sig.) ^t		df	Sig. (2-tailed)	(2-Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
Pre- test	Equal variances assumed	0.026	0.873	3.180	58	0.002	0.9333	0.2935	0.3457	1.5209
	Equal variances not assumed			3.180	57.996	0.002	0.9333	0.2935	0.3457	1.5209

Given the initial difference in the pre-test scores, an Analysis of Covariance (ANCOVA) was conducted to compare the post-test performance while statistically controlling for the pre-test baseline as a covariate, as shown in Table 2. Preliminary checks confirmed that the core assumptions of ANCOVA, including the homogeneity of regression slopes ($F(1, 56) = 1.04$, $p = .312$) and the homogeneity of variances (Levene's test: $F(1, 58) = 2.84$, $p = .097$), were fully satisfied. After adjusting for the covariate, the ANCOVA results demonstrated a statistically significant main effect for the instructional method, $F(1, 57) = 13.39$, $p < .001$, $\eta^2 = 0.190$. The partial eta squared value ($\eta^2 = 0.190$) indicates a large effect size, meaning that the group effect accounted for 19% of the total variance in post-test communication scores. The analysis of adjusted means further clarified this adjusted group difference: when evaluated at an equalized baseline (an overall pre-test mean of 7.17), the experimental group achieved a significantly higher adjusted post-test mean ($M_{adj} = 7.72$, $SE = 0.13$) compared to the control group ($M_{adj} = 7.39$, $SE = 0.13$).

Table 2

ANCOVA Results for Post-test English Communication Performance

	Type III Sum of Squares	df	Mean Square	F	Sig.
Covariate (Pre-test)	35.812	1	35.812	64.912	< .001
Group (Method)	7.388	1	7.388	13.391	< .001
Error	31.448	57	0.552		
Total	3492.250	60			

A Paired Samples t-test was conducted to compare students' pre-test and post-test scores within each group. The results (as shown in Table 3 and Table 4) showed different outcomes between the control and experimental groups. For the control group, the mean score decreased slightly from 7.633 to 7.607, but this difference was not statistically significant, $t(29) = 0.178$, $p = .860$. A notable improvement was observed in the experimental group, with its mean score rising from 6.700 at pre-test to 7.497 at post-test, $t(29) = -7.333$, $p < .001$. These findings indicate that students who used *Speakee* achieved greater gains in English speaking performance during the intervention period. Notably, while the control group showed little change over six weeks, the experimental group demonstrated clear and consistent progress.

Table 3

Paired Sample t-test on Pre- and Post-Intervention Scores of Control and Experimental Groups

Group		Mean	N	Std. Deviation	Std. Error Mean
Control	Pre-test	7.6333	30	1.1418	.2085
	Post-test	7.6067	30	1.0455	.1909
Experimental	Pre-test	6.7000	30	1.1320	.2067
	Post-test	7.4967	30	1.4126	.2579

Table 4

Descriptive Statistics for Pre- and Post-Intervention Scores

Comparison	Mean Diff.	SD	SE	95% CI Lower	95% CI Upper	t	df	p
Control Pre-test - Control Post-test	0.0267	0.8183	0.1494	-0.2789	0.3322	0.178	29	.860
Experimental Pre-test - Experimental Post-test	-0.7967	0.5951	0.1086	-1.0189	-0.5745	-7.333	29	< .001

The descriptive statistics in Table 5 indicated that the post-test scores of the two groups were relatively close. The control group achieved a slightly higher mean score ($M = 7.607$, $SD = 1.045$) than the experimental group ($M = 7.497$, $SD = 1.413$). Levene's test showed that the assumption of equal variances was met ($F = 2.843$, $p = .097$), allowing the use of the equal variances assumed results.

The post-test results showed a smaller difference in mean scores between the two groups than was observed at the pre-test stage. Although the control group retained a slightly higher mean score, the difference was no longer statistically significant. The improvement in mean score was more pronounced among students in the experimental group than among those in the control group.

Table 5

Descriptive Statistics of Post-test Scores for Control and Experimental Groups.

	Group	N	Mean	Std. Deviation	Std. Error Mean
<i>Post-test</i>	0 (Control)	30	7.6067	1.0455	.1909
	1 (Experimental)	30	7.4967	1.4126	.2579

The Independent Samples *t*-test results in Table 6 indicated no significant difference in post-test performance between the two groups, $t(58) = 0.343$, $p = .733$. Levene's test confirmed that the assumption of equal variances was met ($F = 2.843$, $p = .097$).

Table 6Independent Samples *t*-test Results on Post-Intervention Scores.

	Levene's Test for Equality of Variance		T-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	(2-Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	2.843	.097	.343	58	.733	.1100	.3209	-.5323	.7523
Equal variances not assumed			.343	53.438	.733	.1100	.3209	-.5334	.7534

Despite comparable post-test performances, the experimental group had significantly lower pre-test scores. Following the intervention, this difference was no longer statistically significant. To examine the magnitude of the intervention effect, effect sizes were calculated for both the within-group comparisons and the ANCOVA model (Table 7). The paired-samples *t*-test showed a large effect size for the experimental group (Cohen's $d = 1.34$), whereas the control group showed a negligible effect size (Cohen's $d = 0.03$). In addition, the ANCOVA yielded a partial eta squared (η^2) of .190 for the group effect, indicating a large effect size. This value suggests that the group effect accounted for 19.0% of the variance in post-test communication scores after controlling for pre-test differences.

Table 7

Summary of Effect Sizes for Within-Group Changes and Between-Group Intervention Effects

		Statistical Metric	Value	Interpretation (Cohen, 1988)
Within-Group Progression (Pre-test vs. Post-test)				
Experimental Group (n = 30)		Cohen's d	1.34	Large (Highly substantial growth)
Control Group (n = 30)		Cohen's d	0.03	Negligible (No meaningful change)
Between-Group Intervention Comparison (ANCOVA)				
Group Effect Method)	(Instructional Partial Eta Squared)		0.190	Large (19% of variance explained)

Research question 2: What benefits and drawbacks did non-English major EFL students perceive when using *Speakee* for their communication skills development?

Perceived Utility and Pedagogical Effectiveness

The qualitative findings suggested that students perceived *Speakee* as supporting their speaking practice through task-based activities situated in realistic, workplace-oriented scenarios. Several participants reported that practicing in a variety of communicative situations helped them respond more readily during conversations. As one student noted, “*practicing with different situations on Speakee helps me react faster when speaking.*” Others also valued the structured design of the activities, indicating that it helped them organize their responses and monitor their performance. One participant commented that “*the three-task structure guides me step by step and provides clear feedback*”. These findings suggest that the contextualized tasks and structured learning process helped students practice speaking and self-monitor during communication.

Convenience and Flexibility

Participants also perceived the accessibility of *Speakee* as supporting independent learning. Several noted that the platform could be used conveniently across different settings and devices, allowing them to practice whenever opportunities arose. As one student explained, “I can use *Speakee* anywhere, even when I’m at home or waiting somewhere outside.” Another participant emphasized its flexibility by stating that “it’s very convenient because I can practice on different devices like my phone or laptop whenever I have time.” This accessibility appeared to encourage additional speaking practice beyond the classroom, as reflected in another participant’s comment: “I tried to spend more time practicing English on the website after class.”

Confidence and Reduced Anxiety

The qualitative findings suggested that students perceived *Speakee* as providing a supportive environment for speaking practice. Several participants reported that interacting with an AI partner reduced their concerns about making mistakes and allowed them to practice English more comfortably. One participant explained, “I felt more confident when speaking because I learned some useful vocabulary from the website.” Others also indicated that the contextualized

role-play scenarios helped them sustain conversations by providing relevant communicative situations, as some students shared, “the variety of topics helped me reduce the fear of having nothing to say”. Overall, participants perceived the platform as offering a flexible environment that encouraged additional speaking practice beyond the classroom.

Technical Limitations

The qualitative findings also identified several challenges that participants encountered when using *Speakee*. A commonly reported issue was the accuracy of the speech-to-text (STT) function, which occasionally failed to recognize learners’ speech correctly. As one participant explained, “Sometimes I say it correctly, but the system still records it wrong, which makes me a bit confused.” Participants also commented on the feedback provided by the platform. Although several found it useful for pronunciation and vocabulary practice, some said the platform offered limited grammar feedback. One student noted, “Sometimes I don’t clearly see my grammar mistakes, so I don’t know how to improve them.” In addition, a few participants observed that the Dragon Egg gamification feature became less engaging with repeated use. As one student commented, “At first, the Dragon Egg feature was interesting, but after a while, it feels a bit repetitive if there are no new features.” These responses indicate that participants perceived limitations related to speech recognition, grammar feedback, and the long-term appeal of the gamification feature.

Discussion

The findings suggest that *Speakee*-supported speaking practice was associated with improved English speaking performance among non-English majors, while participants also reported generally positive perceptions of the platform. Although the experimental group began the study with significantly lower pre-test scores, ANCOVA showed a statistically significant group effect after controlling for baseline scores. These results suggest that participation in AI-supported speaking activities through *Speakee* was associated with improvements in students’ English communication performance. One possible explanation is that the platform offered additional opportunities for exposure to contextualized speaking practice beyond regular classroom instruction. Through repeated participation in role-play activities and automated feedback, students were able to practice English communication more frequently. This is consistent with Task-Based Language Teaching (Ellis, 2003, 2009), which emphasizes learning through meaningful communicative tasks, and Situated Learning Theory (Lave & Wenger, 1991), emphasizing the crucial role of authentic contexts for language development.

The present findings are broadly consistent with previous studies conducted in international contexts. Consistent with Chen et al. (2025), Du and Daniel (2024), and Fathi et al. (2024), the present study found that AI-supported role-play activities were associated with improvements in learners’ speaking performance. However, unlike Chen et al. (2025), who reported no significant difference in speaking gains between AI-mediated and peer role-play activities, the present study found that the experimental group demonstrated greater improvement after controlling for initial differences. This difference may be related to the design of *Speakee*, which combined AI-supported interaction with contextualized workplace scenarios and structured feedback. Importantly, the users could generate the scenarios that matched their communication needs in real life.

Likewise, the findings are comparable to those reported by Mingyan et al. (2025) that learners using the application Liulishuo achieved higher overall speaking performance than the control group. Although the present study did not examine individual speaking components separately,

interview data suggested that students perceived pronunciation feedback as one of the most beneficial features of the platform, which is broadly consistent with Mingyan et al.'s findings regarding pronunciation improvement. Furthermore, participants' positive perceptions of *Speakee* as a flexible and supportive learning environment correspond with the research by Park et al. (2026), who reported that personalized AI-supported interaction enhanced learner engagement and motivation. While the present study did not directly measure motivational constructs, students' perceptions suggest that contextualized AI-supported practice may encourage greater participation in speaking activities beyond the classroom.

The study's outcomes also extend previous research conducted in Vietnamese contexts. Nguyen and Nguyen (2023) reported that learners valued ELSA Speak for its pronunciation guidance, detailed feedback, and user-friendly interface, although repetitive practice activities reduced learner engagement over time. Similar perceptions emerged in the present study, where participants appreciated *Speakee*'s pronunciation feedback but also identified limitations related to speech recognition accuracy and the long-term appeal of the gamification features. In addition, the present findings support those of Duong and Suppasetserree (2024), who found that the Andy chatbot contributed to improvements in English speaking skills and was positively received by non-English major students. Unlike these commercial AI applications, however, *Speakee* was specifically developed to provide contextualized role-play scenarios aligned with learners' academic and future workplace communication needs. This context-specific design may explain why participants highlighted the relevance of the speaking tasks and the flexibility of practicing beyond the classroom.

Despite these positive findings, several limitations exist. First, the experimental group's significantly lower pre-test scores suggest a higher potential for rapid initial improvement, despite statistical controls. Second, the short six-week intervention may have introduced a novelty-related engagement effect. This aligns with qualitative concerns regarding *Speakee*'s long-term appeal. Third, individual usage varied because practice outside the classroom was self-directed, which may have influenced individual outcomes. Future research should ensure stricter baseline equivalence, conduct longitudinal studies, and track precise user-log data to address these variables.

A critical technical limitation involves Speech-to-Text (STT) accuracy when processing non-native accents, which cascaded into three main issues. First, it compromised feedback accuracy; because the AI generates responses based on transcribed texts, STT errors could trigger misdirected or irrelevant corrections. Second, it hindered the learner experience, as frequent misrecognitions caused communication breakdowns during role-plays, inducing frustration and reducing learners' trust in the platform. Finally, these technical flaws complicate the interpretation of effectiveness. If system interactions are disrupted by transcription errors, the observed speaking gains may not fully capture learners' actual communicative competence. Thus, refining STT models for EFL accents remains essential.

In conclusion, the findings suggest that AI-supported speaking platforms can provide valuable opportunities for contextualized communication practice among non-English major EFL students. The qualitative data emphasize that the effectiveness of such platforms may depend not only on opportunities for interaction but also on the quality of speech recognition, feedback mechanisms, and sustained learner engagement. These aspects should therefore be considered in the continued development of AI-supported language learning platforms.

Conclusion

This research evaluated the extent to which a self-developed website could support non-English major students' English speaking performance. The experimental group showed significant within-group improvement, while ANCOVA indicated a significant adjusted group effect after controlling for baseline scores. Qualitative findings further suggested that students perceived the platform as supporting speaking practice through contextualized role-play tasks, pronunciation feedback, flexible access, and opportunities for additional practice beyond the classroom. Participants also identified several limitations, including speech recognition accuracy, grammar-related feedback, and the long-term appeal of the gamification features.

The findings indicate the potential of Speakee as a supplementary tool for contextualized English speaking practice in higher education. For students, the platform offers additional opportunities for contextualized speaking practice beyond classroom instruction. For teachers, it may be integrated into blended learning environments to increase learners' speaking practice and engagement. However, given the current technical limitations, the platform should be used to complement rather than replace classroom instruction or formal speaking assessment. Unlike commercial speaking-learning tools that mainly provide predetermined content, *Speakee* integrates user-generated role-play scenarios, voice interaction, structured feedback, progress tracking, and gamification tailored to the academic and workplace communication needs of non-English majors.

Several limitations should be acknowledged. First, researchers should consider the initial differences between the control and experimental groups when interpreting the intervention effect, despite using ANCOVA. Second, the six-week intervention may not have been long enough to evaluate longer-term changes in learners' communication development. Third, variations in students' participation throughout the intervention may have influenced the observed learning outcomes. Future studies should investigate whether AI-supported speaking practice can produce sustainable improvement over a longer period of time. Further studies may also explore improvements in speech recognition, grammar feedback, and adaptive gamification features to enhance learners' speaking experiences and sustained engagement.

Declaration of the use of AI

During the preparation of this work, the authors used Gemini to proofread the manuscript, improve readability, and refine academic phrasing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical statement

Informed consent was obtained from all participants before their involvement in the study. Participation was entirely voluntary, and individuals were informed that they could withdraw at any time without penalty. To ensure privacy, all collected data were treated confidentially and used only for research purposes.

Acknowledgements

We thank the participants who took part in the study.

References

- Baek, J., & Lee, C. H. (2018). University students' perceptions and engagement in mobile-assisted blended learning in English speaking classes. *Multimedia-Assisted Language Learning*, 21(4), 11-36. <https://doi.org/10.15702/mall.2018.21.4.11>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 2–27). London: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <http://dx.doi.org/10.1093/applin/I.1.1>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chen, M. R. A., & Hwang, G. J. (2019). Effects of a concept mapping-based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology*, 51(3), 817-834. <https://doi.org/10.1111/bjet.12887>
- Chen, A., Jia, J., Li, Y., & Fu, L. (2025). Investigating the effect of role-play activity with GenAI agent on EFL students' speaking performance. *Journal of Educational Computing Research*, 63(1), 99–125, <https://doi.org/10.1177/07356331241299058>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). <https://doi.org/10.1145/2181037.2181040>
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6, Article 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Duong, T., & Suppasetseree, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning, Teaching and Educational Research*, 23(3), 293–321. <https://doi.org/10.26803/ijlter.23.3.15>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–246. <https://doi.org/10.1111/j.1473-4192.2009.00231.x>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, Article 103254. <https://doi.org/10.1016/j.system.2024.103254>

- Ho, P. X. P. (2024). Using ChatGPT in English language learning: A study on I.T. students' attitudes, habits, and perceptions. *International Journal of TESOL & Education*, 4(1), 55–68. <https://doi.org/10.54855/ijte.24414>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics. Selected Readings* (pp. 269–293). Penguin.
- Kessler, G. (2010). Fluency and anxiety in self-access speaking tasks: The influence of environment. *Computer Assisted Language Learning*, 23(4), 361–375. <https://doi.org/10.1080/09588221.2010.512551>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(3), 537–550. <https://doi.org/10.1177/00336882231162868>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Ngo, M. T., & Tran, L. T. (2023). Current English education in Vietnam: Policy, practices, and challenges. In *English language education for graduate employability in Vietnam* (pp. 49–69). Springer Nature, Singapore. https://doi.org/10.1007/978-981-99-4338-8_3
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/S0958344023000113>
- Nguyen, K. N., & Nguyen, T. M. T. (2023). Non-specialized students' benefits and challenges in using ELSA Speak application for pronunciation learning. *European Journal of Alternative Education Studies*, 8(4), 185–198. <http://dx.doi.org/10.46827/ejae.v8i4.5149>
- Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. *Humanities and Social Sciences Communications*, 12(1), Article 370. <https://doi.org/10.1057/s41599-025-04688-0>
- Park, M., Lee, S., Ma, J., & Yoon, D. (2026). AI Twin: Enhancing ESL speaking practice through AI self-clones of a better me. *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, Article 78, 1–21. <https://doi.org/10.1145/3772318.3790266>
- Pham, T. S. (2025). Using task-based language teaching to improve English speaking fluency: An action research study with Vietnamese non-English-majored university students. *International Journal of TESOL & Education*, 5(4), 16–31. <https://doi.org/10.54855/ijte.25542>
- Rao, P. S. (2019). The importance of speaking skills in English classrooms. *Alford Council of International English & Literature Journal*, 2(2), 6–18.
- Shan, S., Grigoryan, G., Ning, H., & Harutyunyan, H. (2024). AI Chatbots: Developing English Language Proficiency in EFL Classroom. *Arab World English Journal*, April 2024, 292–305. <https://dx.doi.org/10.24093/awej/ChatGPT.20>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation.

- Frontiers in Psychology*, 14, 1210187. <https://doi.org/10.3389/fpsyg.2023.1210187>
- Tai, T.-Y., & Chen, H. H.-J. (2023). 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>
- Trinh, N. B., & Pham, D. T. T. (2021). Challenges in speaking classrooms among non-English majors. *Vietnam Journal of Education*, 5(2), 37–42. <https://doi.org/10.52296/vje.2021.52>
- Vasu, K., Yunus, M. M., & Said, N. E. M. (2025). AI-enhanced interactive platforms for language learning: Examining motivation and proficiency development in primary school students. *International Journal of Academic Research in Progressive Education and Development*, 14(1), 1922–1946, <https://doi.org/10.6007/IJARPED/v14-i1/24851>
- Willis, J. (1996). *A framework for task-based learning*. Longman.

Biodata

Nguyen Le Nguyen and Pham Thi Thao Nguyen have just graduated from the University of Foreign Language Studies, the University of Da Nang, Vietnam. Their research interests focus on the integration of educational technology and AI-driven applications into English language teaching and learning.

Dr Uyen Nu Thuy Nguyen has been teaching at the University of Foreign Language Studies, The University of Da Nang for 18 years. Her academic interest areas are online education, distance learning, e-learning, computer-assisted language learning, and teaching methodology. She has published articles in WoS- and Scopus-indexed journals and in journals in international indexes.