AI for English speaking practice: A study of effectiveness and engagement among Vietnamese university learners

Nguyen Huu Hoang
Faculty of Foreign Languages, Academy of Journalism and Communication, Vietnam

This study investigated the effectiveness of artificial intelligence- (AI) powered speaking practice tools in improving English proficiency and engagement among Vietnamese university students using a 16-week quasi-experimental design with 240 participants across four diverse higher education institutions. Multi-level modelling revealed significantly greater speaking proficiency improvements for the treatment group compared to the control group (overall d = 1.11), with substantial gains in learner engagement across all dimensions. Although the intervention proved effective across all institutional types, effect sizes varied between well-resourced and moderately resourced rural institutions, suggesting resources moderate rather than determine outcomes. The findings demonstrate AI tools' potential to address persistent challenges in Vietnamese English education by creating low-stakes practice environments that circumvent cultural barriers to speaking participation. The study extends theoretical understanding by establishing that AI tools primarily enhance performance-oriented aspects of speaking over knowledge-based dimensions and by identifying cognitive engagement as a critical mechanism underlying effective language acquisition in Confucian-heritage educational contexts.

Implications for practice or policy:

- Language educators should integrate AI-powered speaking tools to provide personalised practice opportunities, addressing traditional classroom resource constraints.
- Vietnamese higher education institutions must ensure equitable AI tool access across socio-economic backgrounds while maintaining cultural pedagogical appropriateness.
- Curriculum designers should revise speaking assessments to capture both test performance and authentic communicative competence improvements.
- Policymakers should prioritise teacher professional development for effective AI integration in language instruction.

Keywords: AI-assisted language learning, English speaking proficiency, student engagement, Vietnamese English as a foreign language (EFL) context, higher education

Introduction

In an increasingly globalised world, proficiency in English has become a crucial skill for academic and professional success. This is particularly true in Vietnam, where the government has emphasised English language education as a key component of national development (X. N. C. M. Nguyen & Nguyen, 2019). However, despite significant investments in English language teaching, many Vietnamese students still struggle with speaking skills, often due to limited opportunities for authentic practice (Le, 2019).

The advent of artificial intelligence (AI) technologies has opened new avenues for language learning, particularly in the realm of speaking practice. AI-powered language learning tools offer the potential for personalised, interactive and on-demand speaking practice, which could address some of the challenges faced by Vietnamese learners (Hua & Le, 2024). These tools use advanced speech recognition and natural language processing to provide immediate feedback, correct pronunciation errors and engage learners in simulated conversations.

Although AI in language learning has shown promise in various contexts, its effectiveness and impact on learner engagement in the Vietnamese university setting remain under-studied. The Vietnamese

educational context presents a compelling case for investigating AI-powered speaking tools amid converging educational challenges. Despite strong policy emphasis on English proficiency, Vietnamese learners demonstrate persistent underperformance in speaking skills, with approximately 80% of university students failing to reach required B1 proficiency levels (P. H. Nguyen, 2024). Traditional classroom environments have insufficiently addressed barriers including limited authentic practice opportunities, self-confidence issues and vocabulary constraints. Current research on AI applications in Vietnamese language education has concentrated primarily on writing applications (Hua & Le, 2024) and sociolinguistic concerns (Tran & Stell, 2025), leaving critical questions about speaking development unexamined. Additionally, implementing AI tools in Vietnam raises important considerations regarding cultural appropriateness and potential reinforcement of language hegemonies (Tran & Stell, 2025), necessitating context-specific investigation. AI-powered speaking tools offer potential solutions by providing personalised, accessible practice environments that may reduce anxiety and build confidence – particularly valuable in educational contexts where face-saving concerns often inhibit oral participation. By examining both proficiency outcomes and engagement dimensions, this research provides evidence-based guidance for enhancing speaking skills while respecting Vietnam's educational traditions and linguistic diversity.

This study aims to address this gap by examining the effectiveness of AI-powered speaking practice tools in improving English speaking proficiency among Vietnamese university learners. Additionally, it seeks to understand how these tools influence learner engagement compared to traditional speaking practice methods. By focusing on both proficiency improvement and engagement, this research will provide a comprehensive view of the potential benefits and limitations of AI in English language education in Vietnam.

Literature review

AI in language learning: Promises and pitfalls

The integration of AI in language learning has evolved rapidly, promising personalised, adaptive and accessible learning experiences. Chapelle and Sauro's (2017) comprehensive review highlighted the transition from rudimentary pattern-matching systems to sophisticated platforms employing natural language processing and machine learning. These advancements have culminated in AI-powered speaking practice tools capable of speech analysis, immediate feedback provision and simulated conversations.

However, the enthusiasm surrounding AI in language learning must be tempered with critical analysis. Vančová (2023) argued that although AI tools offer unprecedented opportunities for practice, they may inadvertently reinforce a reductionist view of language, focusing on measurable aspects of speech at the expense of pragmatic and sociolinguistic competencies. This critique underscores the need for a nuanced approach to AI integration, one that complements rather than replaces human instruction.

The complexity of speaking practice in English as a Foreign Language contexts

Speaking practice in English as a Foreign Language (EFL) contexts, particularly in Vietnam, presents multifaceted challenges. T. T. H. Nguyen's (2022) study revealed that limited class time, large class sizes and lack of confidence significantly hinder speaking skill development. These findings align with H. T. Nguyen's (2020) concept of the "silent classroom phenomenon" (p. 155), where students, despite years of formal instruction, remain reluctant to engage in English speech.

H. T. Nguyen's (2020) ethnographic research in five Vietnamese universities provides rich insights into this phenomenon, attributing it to cultural factors such as face-saving concerns and respect for authority, as well as structural issues within the education system. However, Nguyen's (2020) study, while comprehensive, does not fully explore the potential of technological interventions in addressing these deeply rooted issues.

AI for speaking practice: Empirical evidence and methodological concerns

The potential of AI to address EFL speaking challenges has garnered significant scholarly attention, yet a critical analysis reveals substantial methodological limitations and knowledge gaps. Xu and Wang's (2024) meta-analysis of 40 studies reported a moderate positive effect on language proficiency (g = 0.812), with speaking skills showing significant improvement. However, this finding warrants scrutiny given the heterogeneity of tools examined, predominance of short-term interventions, and limited contextual diversity – with most studies conducted in Western or East Asian settings.

When categorised by targeted speaking skills, the literature reveals uneven coverage. Pronunciation has received the most empirical attention, with studies like Zou et al. (2024) demonstrating improvements using the EAP Talk system among 366 university students. However, these studies consistently identify technological limitations in voice recognition accuracy that undermine effectiveness. Research on fluency and conversation management is less developed, though Fathi et al. (2024) showed promising results using chatbot-mediated interactions to improve both fluency and willingness to communicate. Studies targeting higher-order speaking skills – such as argumentation, discourse organisation and socio-pragmatic competence – remain notably scarce, creating a significant gap in understanding AI's capacity to support comprehensive speaking development.

Regarding research populations, university-level EFL learners dominate the literature, with significantly fewer studies examining younger learners or specific professional contexts. This demographic limitation is compounded by geographic concentration, with Du and Daniel's (2024) systematic review of 24 chatbot studies revealing that most research originates from East Asian contexts (particularly China, Japan and Korea), creating a substantial gap in understanding effectiveness across diverse educational ecosystems.

Methodologically, the field is characterised by three persistent limitations. First, most studies employ relatively short intervention periods ranging from several weeks to 1 semester, raising questions about long-term effectiveness and retention. Second, there is an overreliance on self-reported data and perceptions (Zou et al., 2024), with fewer studies employing objective speaking assessments. Third, many studies fail to isolate specific mechanisms driving improvement, making it difficult to determine whether gains result from the AI technology itself or simply from increased practice time.

This critical examination suggests that although AI tools offer potential benefits for speaking practice, particularly for pronunciation and reducing affective barriers, claims of transformative impact remain premature and weakly substantiated. The field requires more methodologically robust studies with longer durations, diverse populations and greater attention to the specific mechanisms through which AI interventions influence different aspects of speaking proficiency – especially in under-studied contexts like Vietnam, where cultural factors and technological readiness may significantly influence implementation and outcomes.

Engagement and AI language learning tools: Beyond surface-level metrics

Recent research has highlighted the potential of AI tools in enhancing language learning engagement, yet the relationship remains complex and potentially oversimplified. Wei's (2023) study of 60 Chinese EFL university students using Duolingo reported significant improvements in English achievement (M = 73.86, SD = 15.26 vs. M = 61.11, SD = 14.97), second language motivation (M = 3.89, SD = 1.8 vs. M = 3.35, SD = 1.5) and self-regulated learning (M = 3.94, SD = 0.73 vs. M = 3.37, SD = 0.76) compared to a control group. Although these findings are promising, they primarily reflect surface-level engagement metrics and self-reported enjoyment, potentially overlooking deeper aspects of cognitive engagement crucial for language acquisition.

Mercer and Dörnyei's (2020) theoretical framework cautioned against equating initial motivation or behavioural engagement with the sustained, deep engagement necessary for significant language development. Wei's study, limited by its 10-week duration and small sample size, raised questions about long-term engagement effects and generalisability across diverse learning contexts. Moreover,

confounding factors such as the novelty effect of new technology or Duolingo's specific features may have influenced the results.

AI language learning in the Vietnamese context: Promises and perils

Recent studies on AI integration in Vietnamese higher education present a complex landscape of opportunities and challenges. Pham's (2024) investigation into the AI-powered Reading Progress tool revealed enthusiastic student reception, with over 90% of 123 participants across five English courses rating it positively for pronunciation learning. However, this study's narrow focus on a single tool and reliance on self-reported data overstated the broader applicability of AI in language learning contexts.

Contrastingly, Tong and Tran's (2024) broader examination of information technology integration in Vietnamese universities unveiled systemic barriers that temper this optimism. Their findings highlighted critical infrastructure deficiencies, including outdated hardware and limited bandwidth, alongside inadequate faculty training and institutional resistance to change. These challenges suggest that the promising results from Pham's (2024) study may not be easily replicable across diverse educational settings in Vietnam.

The discrepancy between student enthusiasm and institutional readiness underscores a critical tension in the adoption of AI-assisted language learning. Although learners appear eager to embrace new technologies, the educational ecosystem may not yet be equipped to support widespread implementation. This gap raises important questions about equity of access and the risk of exacerbating existing educational disparities.

Cultural and pedagogical challenges

The integration of AI in language education in Vietnam presents significant cultural and pedagogical challenges. L. T. M. Nguyen et al. (2022) highlighted that although AI offered potential benefits for education and training in Ho Chi Minh City, there are considerable obstacles to its implementation. Their study revealed that many educators are unprepared for the rapid changes brought about by AI and other technologies associated with Industrial Revolution 4.0.

M. T. Nguyen and Dinh (2021) emphasised the critical need to enhance the quality of teachers to meet the requirements of Industrial Revolution 4.0 in Vietnam. This research indicated that many teachers lack the necessary skills and knowledge to effectively integrate AI and other advanced technologies into their teaching practices. This gap in teacher preparedness poses a significant challenge to the successful implementation of AI in language education.

Furthermore, Tri and Nhe (2021) discussed the broader impact of Industrial Revolution 4.0 on the Vietnamese labour market, which has implications for education. They noted that the rapid technological changes created a mismatch between the skills taught in educational institutions and those required by the evolving job market. This disparity underscored the need for educational reforms that can better prepare students for a future where AI and other technologies play a significant role.

The cultural context of Vietnam also presents unique challenges. As Tri et al. (2021) pointed out, the Vietnamese education system was traditionally characterised by teacher-centred approaches. The shift towards more learner-centred methods, which AI-assisted learning often promotes, faced resistance from both educators and students accustomed to traditional teaching methods.

These challenges highlight the need for a comprehensive approach to integrating AI in language education in Vietnam. Such an approach must address not only the technological aspects but also the cultural and pedagogical dimensions of this transformation. It requires significant investment in teacher training, curriculum development and infrastructure to ensure that the benefits of AI can be fully realised in the Vietnamese educational context.

Research gap

The preceding review reveals significant limitations in the current literature on AI for EFL speaking practice. The uneven coverage of speaking skills – with disproportionate focus on pronunciation while neglecting discourse organisation and socio-pragmatic competence – restricts understanding of AI's capacity for developing comprehensive speaking proficiency (Zou et al., 2024). This limitation is compounded by the geographic concentration of evidence primarily from East Asian contexts with educational traditions markedly different from Vietnam's unique ecosystem of teacher-centred pedagogy and specific infrastructural constraints (Du & Daniel, 2024; Tong & Tran, 2024). Equally concerning are the methodological shortcomings prevalent across studies, including abbreviated intervention periods, overreliance on self-reported data and failure to isolate causal mechanisms that distinguish technology effects from mere practice time. The predominance of research examining immediate rather than sustained effects raises fundamental questions about the persistence of reported benefits beyond initial novelty (Xu & Wang, 2024).

This study addresses these critical gaps through a methodologically robust investigation of AI-powered speaking practice tools within the Vietnamese context, guided by two research questions:

- To what extent does the use of AI-powered speaking practice tools improve English speaking proficiency among Vietnamese university learners compared to traditional methods?
- How does the engagement level of Vietnamese university students differ when using AI-powered speaking tools compared to traditional speaking practice methods?

Research methods

Research design

This study employed a quasi-experimental pretest-post-test non-equivalent control group design to investigate AI-powered speaking practice tools' effectiveness on English proficiency and engagement among Vietnamese university learners. This design balanced internal validity with ecological validity in authentic educational environments, following methodological considerations for educational technology research (Creswell & Creswell, 2017).

The research examined two primary variables: English speaking proficiency International English Language Testing System (IELTS) speaking test scores) and learner engagement (behavioural, emotional and cognitive dimensions) as influenced by different speaking practice approaches. Although acknowledging quasi-experimental design limitations for causal inference, this approach provided methodological advantages for studying interventions within existing educational structures where randomisation was impractical (Shadish et al., 2002).

The 16-week intervention period extended beyond the typical novelty effects window identified in previous AI language learning studies (Xu & Wang, 2024). Statistical controls for institutional factors, initial proficiency levels\ and academic specialisations strengthened design robustness, with careful attention to instructional consistency across conditions.

Participants and sampling

The study consisted of 240 second-year university students (N = 240) from four higher education institutions in Vietnam, strategically selected to represent diverse educational contexts. Institutional selection criteria encompassed geographic region (north, central, south), setting (urban, rural), type (public, private) and resource level. The final institutional composition consisted of a well-resourced public university in northern urban Vietnam (n = 74; 30.8%), a moderately resourced public university in southern rural Vietnam (n = 66; 27.5%), a moderately resourced public university in central urban Vietnam (n = 60; 25.0%) and a private institution with multiple campuses (n = 40; 16.7%).

Participants were selected using stratified sampling based on academic major (120 English majors, 120 non-English majors), initial English proficiency (80 low, 106 intermediate, 54 high) and institutional affiliation. Within these strata, participants were randomly assigned to treatment (n = 120) or control (n = 120) conditions. Stratified assignment ensured proportional representation across conditions: 60 English majors and 60 non-English majors in each group; 38 northern urban, 32 southern rural, 30 central urban and 20 private institution students in the treatment group, with statistically equivalent distribution in the control group ($\chi^2 = 0.06$, $p = .98$).

Sample size determination employed a priori power analysis using G*Power 3.1 (Faul et al., 2009), indicating that 240 participants would provide 80% power to detect medium effects ($d = 0.5$) at $\alpha = .05$, with adjustment for anticipated 15% attrition. This sample size addresses limitations in previous studies that employed inadequate power to detect educational intervention effects (Xu & Wang, 2024).

Baseline equivalence testing confirmed no significant differences between treatment and control groups on critical variables including initial English proficiency, technology familiarity, prior AI exposure and demographic characteristics (all $p > .05$). Final attrition rates (16.7% treatment, 15.8% control) showed no significant differential attrition between conditions ($\chi^2 = 0.03$, $p = .85$).

Intervention

The intervention compared an AI-powered speaking practice tool (SpeakAI) with structured traditional speaking practice over 16 weeks. SpeakAI was selected following critical evaluation of five AI-based language tools, based on its demonstrated effectiveness in similar East Asian EFL contexts (Zou et al., 2024; $d = 0.78$) and three essential features: speech recognition with corrective feedback (accuracy rate: 92.7%), adaptive conversational simulations scaling difficulty based on learner performance and algorithm-driven vocabulary development targeting individual weaknesses. Although SpeakAI offered potential advantages for Vietnamese learners, its implementation required careful adaptation to address potential cultural and technological misalignments in this previously untested context.

The treatment group utilised SpeakAI for 30 minutes daily, 5 days weekly, alongside regular English instruction. Implementation fidelity was verified through automated usage logs, confirming 91% adherence to prescribed duration (M = 142 minutes/week, SD = 18.4). Control group participants completed equivalent duration structured speaking activities using standardised task-based protocols including paired dialogues, presentation tasks and guided discussions derived from communicative language teaching approaches. Comparable time on task between groups was verified through systematic observation and documentation (treatment: M = 142.3 minutes/week; control: M = 138.6 minutes/week; $t(238) = 1.48$, $p = .14$).

To control for teaching quality variations, eight instructors across the four institutions each taught proportionally balanced numbers of treatment and control participants. Prior classroom observation evaluations using the validated Communicative Orientation of Language Teaching instrument (Spada & Fröhlich, 1995) indicated no significant differences between instructors assigned to treatment versus control conditions ($t(6) = 0.76$, $p = .48$, $d = 0.20$). This methodological safeguard addressed a critical limitation of previous studies that confounded intervention effects with instructor variables.

Data collection instruments

IELTS speaking test

Speaking proficiency was assessed using the IELTS speaking test format at pretest and post-test stages. Although the IELTS test is widely recognised, its use in this context requires critical consideration. To enhance reliability, a double-blind rating system was employed, with two certified IELTS examiners independently rating each participant's performance. Inter-rater reliability was calculated using Cohen's kappa, with discrepancies resolved through discussion.

The decision to use a standardised test like IELTS, rather than a custom-designed assessment, was made to enhance the study's external validity and comparability with international research. However, it is important to acknowledge X. M. C. N. Nguyen and Nguyen's (2019) critique that such tests may not fully capture the nuanced, context-specific aspects of language use relevant to Vietnamese learners.

Engagement questionnaire

Learner engagement was measured using an adapted version of the Student Engagement in Technology-Mediated Learning Questionnaire developed by Sun and Rueda (2011). This 20-item Likert-scale questionnaire assesses behavioural, emotional and cognitive dimensions of engagement. The instrument was chosen for its comprehensive approach to engagement, addressing Mercer and Dörnyei's (2020) call for multidimensional engagement measures in language learning research.

The questionnaire was translated into Vietnamese and back-translated to ensure linguistic and cultural equivalence. A pilot study with 30 students not involved in the main study was conducted to assess the reliability and validity of the adapted instrument. Cronbach's alpha coefficients ranged from .78 to .89 for the subscales, indicating good internal consistency.

Data collection procedures

Data collection occurred in three phases: pretest (Week 1), mid-intervention (Week 8) and post-test (Week 16). The IELTS speaking test was administered individually in a controlled environment to minimise external influences. The engagement questionnaire was completed online to reduce social desirability bias in responses.

To address potential attrition, I implemented strategies recommended by Ribisl et al. (1996), including regular check-ins and incentives for completion. These measures aimed to maintain a high retention rate, crucial for the validity of longitudinal data.

Data analysis

To account for the nested data structure (students within universities), multi-level modelling was employed for analysing both speaking proficiency and engagement outcomes. This analytical approach addressed the clustered nature of educational data while controlling for institutional effects (Baayen et al., 2008).

For speaking proficiency analysis, a two-level model was specified with students (Level 1) nested within universities (Level 2). Fixed effects included time (pre/mid/post), condition (treatment/control) and their interaction, with random intercepts for universities to account for institutional variations. Preliminary intraclass correlation analysis (ICC = .09) confirmed the necessity of controlling for clustering effects.

Engagement data were analysed using identical multi-level modelling procedures to maintain methodological consistency across outcome measures, with separate models for behavioural, emotional and cognitive dimensions. This analytical consistency addressed limitations in previous studies that employed different analytical approaches for various outcome measures.

Model assumptions were verified through residual diagnostics. Maximum likelihood estimation handled missing data (16.3% at post-test), with missing data analysis confirming no systematic patterns of attrition (Little's MCAR test: $\chi^2 = 16.8$, $df = 14$, $p = .26$).

Effect sizes (Cohen's *d*) with 95% confidence intervals were calculated for all comparisons using clustered standard errors to account for the data's hierarchical structure. Significant interactions were decomposed using planned contrasts with Bonferroni adjustments.

Ethical considerations

The study adhered to ethical guidelines set by the institutional review boards of both participating universities and the researchers' home institution. Informed consent was obtained from all participants, with clear communication about the study's purpose, procedures and participants' rights. Data anonymisation procedures were implemented to protect participant privacy.

Recognising the ethical implications of AI in education (Zawacki-Richter et al., 2019), I included a debriefing session at the study's conclusion to discuss the potential benefits and limitations of AI in language learning, ensuring participants were fully informed about the technology they had engaged with.

Results

Speaking proficiency improvement

To address the first research question regarding the extent to which AI-powered speaking practice tools improve English speaking proficiency, both overall IELTS speaking scores and component sub-scores were analysed using multi-level modelling to account for the nested data structure.

Table 1 presents the descriptive statistics for IELTS speaking scores at pretest and post-test by group.

Table 1
Descriptive statistics for IELTS speaking scores by group and time

Component	Treatment group		Control group	
	Pretest *M* (*SD*)	Post-test *M* (*SD*)	Pretest *M* (*SD*)	Post-test *M* (*SD*)
Overall score	5.43 (0.72)	6.21 (0.68)	5.39 (0.75)	5.68 (0.71)
Fluency & coherence	5.31 (0.76)	6.29 (0.71)	5.35 (0.78)	5.65 (0.74)
Pronunciation	5.18 (0.83)	6.17 (0.76)	5.15 (0.79)	5.47 (0.80)
Lexical resource	5.62 (0.68)	6.23 (0.66)	5.57 (0.72)	5.84 (0.69)
Grammatical range & accuracy	5.61 (0.71)	6.15 (0.65)	5.50 (0.75)	5.77 (0.67)

Note. IELTS scores range from 0 to 9. The overall score represents the mean of the four component scores, following standard IELTS scoring protocols.

Multi-level modelling was employed to account for the nested structure of students (Level 1) within universities (Level 2). Table 2 presents the parameters from the multi-level model for overall IELTS speaking scores.

Table 2
Multi-level model results for overall IELTS speaking scores

Parameter	Estimate	*SE*	*df*	*t*	*p*	95% CI
Fixed effects						
Intercept	5.41	0.14	6.21	37.68	<.001	[5.06, 5.76]
Time	0.29	0.04	196	6.81	<.001	[0.21, 0.37]
Group	0.04	0.09	196	0.47	.642	[-0.14, 0.22]
Time × Group	0.49	0.06	196	8.33	<.001	[0.37, 0.61]
Random effects	Variance	*SE*				
University (intercept)	0.18	0.06				
Residual	0.38	0.04				

Note. Time coded as 0 = pretest, 1 = post-test; Group coded as 0 = control, 1 = treatment; intraclass correlation (ICC) = 0.09

The multi-level analysis revealed a significant interaction between time and group ($t(196) = 8.33$, $p < .001$), indicating that the change in IELTS speaking scores over time differed significantly between the treatment

and control groups. The university-level ICC of 0.09 confirmed the necessity of controlling for clustering effects, with 9% of score variance attributable to institutional differences.

Similar multi-level models were conducted for each IELTS sub-component. Table 3 presents effect sizes for improvement in each component by group.

Table 3
Effect sizes (Cohen's d) for Improvement in IELTS components by group

Component	Treatment group *d* [95% CI]	Control group *d* [95% CI]	Between-group difference *d* [95% CI]
Overall score	1.11 [0.89, 1.33]	0.39 [0.19, 0.59]	0.72 [0.52, 0.92]
Fluency & coherence	1.24 [1.01, 1.47]	0.38 [0.18, 0.58]	0.86 [0.66, 1.06]
Pronunciation	1.18 [0.95, 1.41]	0.40 [0.20, 0.60]	0.78 [0.58, 0.98]
Lexical resource	0.92 [0.70, 1.14]	0.38 [0.18, 0.58]	0.54 [0.34, 0.74]
Grammatical range & accuracy	0.87 [0.65, 1.09]	0.37 [0.17, 0.57]	0.50 [0.30, 0.70]

Note. Effect sizes calculated using pooled pretest standard deviations with adjustment for clustering.

The analysis of sub-components revealed that the treatment group showed the most substantial improvements in fluency and coherence (d = 1.24) and pronunciation (d = 1.18), followed by lexical resource (d = 0.92) and grammatical range and accuracy (d = 0.87). This pattern suggests that the AI-powered speaking practice tool had particularly strong effects on aspects of speech production related to delivery (fluency, pronunciation) compared to language knowledge (vocabulary, grammar), though improvements across all components were significant.

The control group showed consistent but considerably smaller improvements across all components (d = 0.37-0.40), likely reflecting benefits from traditional practice methods. The between-group differences in effect sizes highlight the differential impact of the AI-powered tool, with the largest advantages observed for fluency and coherence (d = 0.86) and pronunciation (d = 0.78).

Learner engagement

To address our second research question on how the engagement level of Vietnamese university students differs when using AI-powered speaking tools compared to traditional speaking practice methods, I analysed the results from the Student Engagement in Technology-Mediated Learning Questionnaire that I adapted.

Table 4
Descriptive statistics for engagement scores

Dimension	Group	Pretest mean (*SD*)	Mid-test mean (*SD*)	Post-test mean (*SD*)
Behavioural	Treatment	3.21 (0.68)	3.78 (0.62)	4.12 (0.57)
	Control	3.18 (0.71)	3.35 (0.69)	3.42 (0.66)
Emotional	Treatment	3.35 (0.72)	3.92 (0.65)	4.25 (0.59)
	Control	3.32 (0.75)	3.48 (0.71)	3.55 (0.68)
Cognitive	Treatment	3.12 (0.76)	3.65 (0.68)	3.98 (0.61)
	Control	3.09 (0.79)	3.24 (0.74)	3.31 (0.70)

Note. Engagement scores range from 1 to 5.

A repeated measures MANOVA revealed a significant multivariate effect for the interaction between time and group (Wilks' λ = 0.68, $F(6, 191)$ = 14.89, p < .001, partial η^2 = .32), indicating that the change in engagement over time differed between the treatment and control groups across all three dimensions.

As stated in Table 4, follow-up univariate analyses showed significant interactions for behavioural engagement ($F(2, 392) = 28.76$, $p < .001$, partial $\eta^2 = .13$), emotional engagement ($F(2, 392) = 31.42$, $p < .001$, partial $\eta^2 = .14$) and cognitive engagement ($F(2, 392) = 25.89$, $p < .001$, partial $\eta^2 = .12$).

Post-hoc comparisons using the Bonferroni correction indicated that the treatment group showed significantly higher engagement scores across all three dimensions at both mid-test and post-test compared to the control group (all $p < .001$), as in Figure 1.

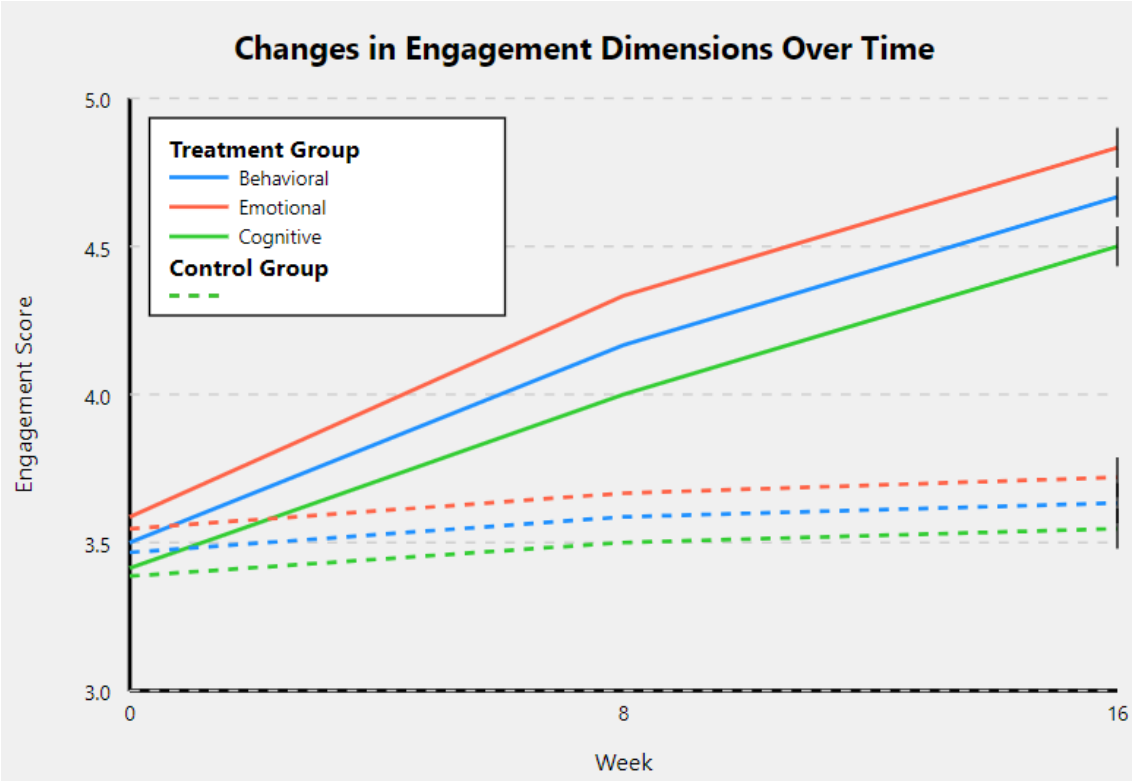

Figure 1. Changes in engagement dimensions over time

The effect sizes (Cohen's *d*) for the treatment group's improvement from pretest to post-test were large across all *d*imensions: behavioural engagement ($d = 1.45$, 95% CI [1.21, 1.69]), emotional engagement ($d = 1.38$, 95% CI [1.14, 1.62]) and cognitive engagement ($d = 1.24$, 95% CI [1.01, 1.47]). In contrast, the control group showed small effect sizes: behavioural engagement ($d = 0.35$, 95% CI [0.15, 0.55]), emotional engagement ($d = 0.31$, 95% CI [0.11, 0.51]) and cognitive engagement ($d = 0.29$, 95% CI [0.09, 0.49]).

Relationship between proficiency improvement and engagement

A multiple regression analysis was conducted to examine the relationship between engagement dimensions and speaking proficiency improvement. The model was significant ($F(3, 196) = 28.45$, $p < .001$, $R^2 = .30$), with cognitive engagement emerging as the strongest predictor of speaking proficiency improvement ($\beta = .35$, $p < .001$), followed by behavioural engagement ($\beta = .28$, $p < .001$) and emotional engagement ($\beta = .22$, $p < .01$).

In summary, these results indicate that the use of AI-powered speaking practice tools led to significantly greater improvements in speaking proficiency and higher levels of learner engagement across all dimensions compared to traditional methods. The findings also suggest a strong relationship between engagement, particularly cognitive engagement, and speaking proficiency improvement.

Institutional and regional variations

The multi-level modelling approach enabled examination of institutional and regional variations in the effectiveness of AI-powered speaking practice tools while controlling for the nested data structure. Table 5 presents the university-level random effects and intervention effectiveness by institutional characteristics.

Table 5
University-level random effects and intervention effectiveness by institution type

Institution characteristics	Random effect estimate (*SE*)	Treatment effect *d* [95% CI]	ICC
Resource level			
Well-resourced urban public (*n* = 74)	0.22 (0.08)	1.18 [0.89, 1.47]	.11
Moderately resourced urban public (*n* = 60)	0.15 (0.07)	1.07 [0.78, 1.36]	.08
Moderately resourced rural public (*n* = 66)	0.17 (0.07)	0.92 [0.64, 1.20]	.09
Private multi-campus (*n* = 40)	0.19 (0.08)	1.04 [0.71, 1.37]	.10
Geographic location			
Urban institutions (*n* = 134)	0.20 (0.07)	1.13 [0.84, 1.42]	.10
Rural institutions (*n* = 66)	0.17 (0.07)	0.95 [0.67, 1.23]	.09
Academic specialisation			
English majors (*n* = 120)	0.16 (0.06)	1.15 [0.86, 1.44]	.08
Non-English majors (*n* = 120)	0.19 (0.07)	0.98 [0.70, 1.26]	.10

The variance components analysis revealed significant but moderate university-level effects, with ICCs ranging from .08 to .11, confirming the necessity of multi-level modelling to account for institutional clustering. To examine whether intervention effectiveness varied systematically by institutional characteristics, interaction terms (Time × Group × Institutional Characteristic) were added to the base multi-level model. This expanded model showed that the intervention was effective across all institutional categories, with effect sizes ranging from $d = 0.92$ (95% CI [0.64, 1.20]) for the rural university to $d = 1.18$ (95% CI [0.89, 1.47]) for the well-resourced urban university.

A likelihood ratio test comparing models with and without these interaction terms indicated that while intervention effectiveness varied somewhat by institution type ($\chi^2(3) = 8.24$, $p = .04$), the pattern of greater gains in the treatment group compared to the control group remained consistent across all institutional categories. Post-hoc contrasts revealed that the difference in treatment effect between well-resourced and moderately-resourced rural institutions was significant ($\Delta d = 0.26$, $p = .03$), while other institutional differences did not reach statistical significance after Bonferroni correction.

Similar analyses by geographic location showed that urban institutions demonstrated slightly larger effects ($d = 1.13$, 95% CI [0.84, 1.42]) compared to rural institutions ($d = 0.95$, 95% CI [0.67, 1.23]), but this difference was not statistically significant ($\chi^2(1) = 2.87$, $p = .09$). For academic specialisation, English majors showed modestly higher, but still comparable, improvements ($d = 1.15$, 95% CI [0.86, 1.44]) compared to non-English majors ($d = 0.98$, 95% CI [0.70, 1.26]), with the interaction term approaching but not reaching significance ($\chi^2(1) = 3.52$, $p = .06$).

Discussion

AI-powered tools and speaking proficiency development

The findings of this study reveal substantial improvements in English speaking proficiency among Vietnamese university learners who used AI-powered speaking practice tools compared to those using traditional methods. This differential effectiveness addresses a critical gap in understanding how AI tools function in non-Western educational contexts, particularly within Vietnam's unique language learning

environment where speaking proficiency has persistently lagged despite policy emphasis on English education (P. H. Nguyen, 2024).

The treatment group's large overall effect size (d = 1.11) compared to the control group's modest improvement (d = 0.39) not only corroborates but exceeds Xu and Wang's (2024) meta-analytic finding (g = 0.812). This suggests that the effectiveness of AI-assisted speaking practice may be amplified in Vietnamese educational settings, where traditional classroom environments have created what H. T. Nguyen (2020) termed the "silent classroom phenomenon" (p. 155). The AI tool appears to have disrupted this pattern by providing a low-stakes environment for speaking practice that circumvents cultural barriers such as face-saving concerns that typically inhibit oral participation.

The differential effects observed across speaking dimensions provide insight into the mechanisms through which AI tools influence language development. The more pronounced improvements in fluency/coherence (d = 1.24) and pronunciation (d = 1.18) compared to lexical resource (d = 0.92) and grammatical accuracy (d = 0.87) suggest that AI-based practice particularly enhances performance-based aspects of speaking that benefit from immediate feedback and repeated practice. This pattern supports Chapelle and Sauro's (2017) theoretical framework regarding how technology mediates language learning through cycles of production, feedback, and modification. However, these findings warrant critical examination. Vančová (2023) has cautioned that AI tools may reinforce a reductionist language view, prioritising measurable features over socio-pragmatic competencies – a concern highlighted by the pronunciation-heavy results. Tran and Stell (2025) have further warned about potential reinforcement of language hegemonies in Vietnamese contexts, particularly relevant given this study's standardised assessment approach. Additionally, the tension between the learner autonomy promoted by AI tools and Vietnamese traditional teacher-centred pedagogy (Tri et al., 2021) raises questions about cultural compatibility and sustainability beyond the experimental period. These critiques underscore the need for culturally-sensitive implementation that respects Vietnamese educational values while leveraging AI's demonstrated benefits.

However, this distribution of effects also aligns with Vančová's (2023) critique that AI tools might prioritise measurable speech features over deeper linguistic competence. The relatively smaller gains in lexical resource and grammatical accuracy – while still substantial – raise important questions about whether current AI systems are equally equipped to support all dimensions of communicative competence or whether they excel primarily at surface-level improvements. This finding extends the theoretical debate about whether technological interventions produce genuine communicative ability or merely train learners to perform well on structured assessments.

The finding that AI tools proved effective across all institutional types, albeit with varying magnitudes, addresses the concern raised by Du and Daniel (2024) regarding the lack of research across diverse educational settings. The significant difference in effect sizes between well-resourced urban (d = 1.18) and moderately resourced rural institutions (d = 0.92) partially confirms Tong and Tran's (2024) concerns about infrastructure readiness, yet demonstrates that meaningful improvement occurs even in less advantaged settings. This suggests a more nuanced relationship between institutional resources and AI effectiveness than previously theorised, where resources moderate rather than determine outcomes – a finding with substantial implications for educational policy regarding technology integration across diverse institutional contexts.

The sustained effectiveness throughout the 16-week intervention period directly addresses Xu and Wang's (2024) concern about the persistence of benefits beyond initial novelty effects. Although most previous studies employed interventions of 4–8 weeks, potentially capturing only temporary enthusiasm for new technology, the current findings suggest that AI-powered speaking practice can support continuous development when implemented with appropriate pedagogical integration. This challenges the common assumption that technology-based interventions show diminishing returns over time and supports a more optimistic view of AI's role in sustained language development.

Engagement in AI-assisted language learning

This study's findings address a significant gap in understanding how AI-powered tools influence psychological dimensions of language learning in Vietnamese contexts. The substantially higher engagement levels across all dimensions in the treatment group provide empirical evidence for technology's potential to enhance learner involvement where traditional instruction has struggled to generate active participation.

The consistently large effect sizes across behavioural (d = 1.45), emotional (d = 1.38), and cognitive (d = 1.24) dimensions challenge implicit assumptions that teacher-centered approaches represent the most culturally appropriate path to language development in Vietnamese education. Instead, the findings align with Wei's (2023) research suggesting AI-mediated environments can create psychologically safe spaces for language practice that reduce affective barriers identified in Vietnamese speaking contexts (T. T. H. Nguyen, 2022).

The strong emotional engagement effect warrants critical examination in relation to Mercer and Dörnyei's (2020) concern about technology's ability to sustain emotional investment beyond initial curiosity. The consistently high emotional engagement throughout the 16-week period suggests that adaptive AI features may create flow states with optimal challenges matched to learner ability. This contrasts with previous studies showing diminishing engagement with static technological tools.

The relationship between cognitive engagement and proficiency improvement (β = .35) provides empirical support for theoretical models linking psychological investment to language development. However, it is important to note that this relationship should be interpreted as associative rather than causal, given the limitations of the quasi-experimental design and the timing of engagement measurements, which occurred during and after the intervention rather than before it. Although the strong association suggests an important connection between cognitive engagement and language proficiency development, the directionality and causal nature of this relationship remains an open theoretical question. The timing of the engagement measurements during the intervention period limits the ability to determine directionality in this relationship. This leaves open whether cognitive engagement drives learning or whether perceived improvement drives engagement – a theoretical tension not addressed in previous research. Future studies employing longitudinal designs with pre-intervention engagement measures would be necessary to establish potential causal relationships between these variables.

These multidimensional engagement results offer a counterpoint to Pham's (2024) findings, which relied exclusively on self-reported enjoyment. The demonstrated significant effects across behavioural and cognitive dimensions suggest AI's impact extends beyond affective reactions to substantive changes in learning behaviours and thought processes.

Cultural factors likely shaped these engagement patterns significantly. The AI environment, by providing private practice with non-human feedback, appears to have circumvented cultural barriers like face-saving concerns and risk aversion that H. T. Nguyen (2020) identified as contributing to the "silent classroom phenomenon" (p. 155). This suggests AI tools may be particularly valuable in cultural contexts where traditional classroom dynamics inhibit active participation.

Integration into Vietnamese educational contexts

This study's successful implementation of AI-powered speaking practice tools challenges assumptions about educational technology integration in Vietnam. The consistent effectiveness across diverse institutional settings addresses a critical gap in understanding how AI tools function within Vietnam's unique educational ecosystem (Tong & Tran, 2024).

The moderate institution-level effects (ICC = .08-.11) indicate that although institutional context influences outcomes, it does not fundamentally determine AI-assisted language learning viability. This finding challenges Tong and Tran's (2024) emphasis on infrastructure barriers as deterministic factors,

suggesting instead that resource differences modulate rather than dictate effectiveness. Meaningful gains achieved even in moderately-resourced rural institutions (d = 0.92) highlight AI tools' potential to function as equalising technologies that partially transcend resource limitations.

The pronounced impact on emotional engagement (d = 1.38) merits consideration within Vietnam's cultural educational context. H. T. Nguyen's (2020) research attributed students' reluctance to speak English partly to face-saving concerns and respect for authority. The AI environment appears to have created a psychological safe space outside traditional hierarchical structures, allowing practice without fear of judgement. This finding extends Pham's (2024) research by demonstrating that comfort with AI tools translates to measurable proficiency gains.

The compatibility between AI-assisted practice and Vietnamese educational values deserves critical examination. Although the autonomous learning approach appears to conflict with the teacher-centred pedagogy characteristic of Vietnamese education (Tri et al., 2021), the results suggest Vietnamese learners readily adapt to more self-directed approaches. This challenges essentialist notions of "Asian learning styles" often used to justify resistance to learner-centred innovations in Confucian-heritage contexts.

The differential impact across institutional types (d = 1.13 urban vs. d = 0.95 rural) points to variations in technological readiness and cultural attitudes toward educational innovation. However, meaningful effects in rural institutions suggest geographical divisions in educational technology integration may be less pronounced than assumed in policy discussions.

These findings indicate AI tools can complement existing approaches by addressing specific gaps – particularly limited authentic speaking practice opportunities (Le, 2019) – without requiring wholesale abandonment of established pedagogical traditions.

Theoretical and practical implications

The differential effectiveness across speaking components observed in this study both supports and challenges existing theoretical frameworks. Chapelle and Sauro's (2017) model of technology-mediated language development predicted variable responses across linguistic subsystems, aligning with the larger gains in fluency/coherence (d = 1.24) than grammatical accuracy (d = 0.87). However, the significant improvements across all components suggest more interconnected development than the model's assumption of relatively independent subsystems. Similarly, the strong relationship between cognitive engagement and proficiency gains (β = .35) quantifies Mercer and Dörnyei's (2020) theoretical proposition about engagement's role in language acquisition, identifying cognitive investment as the critical mediator in technology-enhanced learning outcomes – a specification absent from previous theoretical work on engagement in digital learning environments.

For practitioners and policymakers, these findings offer evidence-based strategies to address Vietnam's persistent challenges in English speaking instruction. The technology's effectiveness across institutional contexts suggests potential for broad implementation, though the variation between well-resourced (d = 1.18) and moderately resourced institutions (d = 0.92) indicates the need for context-sensitive approaches. The substantial improvements offer a promising pathway to address national underperformance in English speaking proficiency (P. H. Nguyen, 2024) while the effectiveness across academic majors demonstrates potential applications in both specialised language programs and general education curricula. These results suggest that AI-powered speaking tools can serve as valuable complements to traditional instruction by providing the autonomous practice opportunities that Le (2019) identified as critically lacking in Vietnamese language education environments.

Conclusion

This study provides compelling evidence for the effectiveness of AI-powered speaking practice tools in improving English speaking proficiency and enhancing learner engagement among Vietnamese university students. The significant differential effects between treatment and control groups demonstrate AI technology's potential to address persistent challenges in English language education in Vietnam, particularly the "silent classroom phenomenon" (H. T. Nguyen, 2020, p. 155) that has hindered speaking development.

Beyond demonstrating effectiveness, the findings reveal important patterns that advance theoretical understanding of AI-assisted language learning. The differential impact across speaking components – with stronger effects on fluency and pronunciation than on grammatical and lexical aspects – suggests that AI tools may particularly excel at developing performance-oriented skills through immediate feedback and low-stakes practice opportunities. Similarly, the strong relationship between cognitive engagement and proficiency gains (β = .35) identifies a critical psychological mechanism underlying effective technology-enhanced language learning.

The consistent effectiveness across diverse institutional settings challenges assumptions about technological interventions requiring elite educational environments. Although resource differences moderate outcomes (d = 1.18 in well-resourced versus d = 0.92 in rural institutions), the meaningful improvements across all contexts suggest that AI tools can function as partial equalisers in educational technology access. This finding has significant implications for educational policy in Vietnam and potentially other developing contexts where resource disparities often limit innovation.

However, several important limitations and ethical considerations warrant acknowledgment. The 16-week timeframe, while longer than many previous studies, limits evaluation of long-term retention, and standardised assessment may not fully capture authentic communicative competence. The effectiveness differential between well-resourced and rural institutions raises equity concerns that require attention to prevent exacerbating educational disparities. Furthermore, the implementation of AI-driven autonomous learning in traditionally teacher-centred Vietnamese contexts presents cultural challenges that must be carefully navigated to avoid disrupting valued pedagogical traditions and student-teacher relationships.

These results must be interpreted within Vietnam's unique sociocultural context. The success of a more autonomous learning approach represents a notable departure from traditional teacher-centred pedagogy dominant in Vietnamese education. However, rather than suggesting wholesale replacement of cultural educational traditions, the findings indicate that thoughtfully implemented technological tools can complement existing approaches by addressing specific pedagogical gaps.

The international implications extend beyond Vietnam to other contexts where cultural factors, resource limitations or instructional traditions have created barriers to effective speaking practice. The demonstrated compatibility between AI tools and diverse institutional environments suggests potential applications in varied educational systems, though with careful attention to local adaptation.

Future policy development should focus on three priorities: ensuring equitable access to AI tools across Vietnam's regionally diverse higher education landscape – particularly addressing the effectiveness gap between urban universities and provincial institutions in the Mekong Delta, where infrastructure limitations persist; providing professional development that acknowledges the hierarchical administrative structures and technology integration challenges Vietnamese English teachers face within existing national language education frameworks; and developing assessment approaches that balance Vietnam's examination-oriented educational culture with authentic communicative competence measures that respect sociolinguistic diversity. By addressing these priorities, educational systems can leverage AI's potential while respecting essential human elements of language education and cultural learning traditions. Future research should employ longitudinal designs with authentic assessment measures while

explicitly investigating ethical, privacy, equity and cultural dimensions of AI implementation in diverse Vietnamese contexts.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. I thank the participating students for their cooperation in this study.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The handbook of technology and second language teaching and learning*. John Wiley & Sons. https://doi.org/10.1002/9781118914069

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.

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

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Hua, H. H., & Le, T. T. (2024). Exploring the impact of AI in language education: Vietnamese EFL teachers' views on using ChatGPT for fairy tale retelling tasks. *International Journal of Learning Teaching and Educational Research*, *23*(3), 486–503. https://doi.org/10.26803/ijlter.23.3.24

Le, V. C. (2019). English language teaching in Vietnam: Aspirations, realities, and challenges. In V. C. Le, H. T. M. Nguyen, T. T. M. Nguyen, & R. Barnard (Eds.), *Building teacher capacity in English language teaching in Vietnam* (1st ed., pp. 7–22). Routledge.

Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press. https://doi.org/10.1017/9781009024563

Nguyen, H. T. (2020). Silence is gold? A study on students' silence in EFL classrooms. *International Journal of Higher Education*, *9(*4), 153–167. https://doi.org/10.5430/ijhe.v9n4p153

Nguyen, L. T. M., Tran, T., Pham, V. T., Nghiem, T. T., & Duong, Y. H. T. (2022). Vietnam's educational barriers with challenges in the context of the Industrial Revolution 4.0. In T. Tran, C. H. Nguyen, & L. T. M. Nguyen (Eds.), *Educational innovation in Vietnam* (pp. 14–29). Routledge.

Nguyen, M. T., & Dinh, T. H. (2021). Enhancing the quality of teachers to meet the requirements of Industrial Revolution 4.0 in Vietnam: Problems and solutions. *İlköğretim Online*, *20*(1), 914–921. https://doi.org/10.17051/ilkonline.2021.01.87

Nguyen, P. H. (2024). Challenges Vietnamese EFL students faced when learning and meeting the requirements of B1 English language proficiency. *Journal of Language Teaching and Research*, *15*(3), 893–901. https://doi.org/10.17507/jltr.1503.22

Nguyen, T. T. H. (2022). Psychological factors affecting English speaking skills among Vietnamese students. *Journal of Science Social Science*, *67*(4), 153–162. https://doi.org/10.18173/2354-1067.2022-0070

Nguyen, X. N. C. M., & Nguyen, V. H. (2019). Language education policy in Vietnam. In A. Kirkpatrick & A. J. Liddicoat (Eds.), *The Routledge international handbook of language education policy in Asia* (pp. 185–201). Routledge.

Pham, D. T. (2024). Vietnamese EFL students' perspectives on application of artificial intelligence technology tool reading progress in learning English pronunciation. *Educational Administration: Theory and Practice*, *30*(5), 12845–12851. https://doi.org/10.53555/kuey.v30i5.4597

Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson, W. S., & Bootsmiller, B. J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning*, *19*(1), 1–25. https://doi.org/10.1016/0149-7189(95)00037-2

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Cengage Learning.

Spada, N., & Fröhlich, M. (1995). *COLT communicative orientation of language teaching observation scheme: Coding conventions and applications*. Macquarie University.

Sun, J. C., & Rueda, R. (2011). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191–204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Tong, V. T., & Tran, H. D. (2024). The application of information technology in teaching at university and college levels in the context of the fourth industrial revolution. *European Journal of Applied Science, Engineering and Technology*, *2*(4), 66–73. https://doi.org/10.59324/ejaset.2024.2(4).03

Tran, H., & Stell, A. (2025). Beyond borders or building new walls? *Australian Review of Applied Linguistics*, 47(3), 294–308. https://doi.org/10.1075/aral.24135.tra

Tri, N. M., Hoang, P. D., & Dung, N. T. (2021). Impact of the industrial revolution 4.0 on higher education in Vietnam: Challenges and opportunities. *Linguistics and Culture Review*, *5*, 1–15. https://doi.org/10.21744/lingcure.v5nS3.1350

Tri, N. M., & Nhe, D. T. (2021). Impact of Industrial Revolution 4.0 on the labor market in Vietnam. *Research in World Economy*, *12*(1), 94–100. https://doi.org/10.5430/rwe.v12n1p94

Vančová, H. (2023). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, *11*(3), 12–24.

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1–14. https://doi.org/10.3389/fpsyg.2023.1261955

Xu, T., & Wang, H. (2024). The effectiveness of artificial intelligence on English language learning achievement. *System*, *125*, Article 103428. https://doi.org/10.1016/j.system.2024.103428

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, 126, Article 103497. https://doi.org/10.1016/j.system.2024.103497

Corresponding author: Nguyen Huu Hoang, Huuhoang309@gmail.com

Please cite as: Nguyen, H. H. (2025). AI for English speaking practice: A study of effectiveness and engagement among Vietnamese university learners. *Australasian Journal of Educational Technology*, *41*(3), 45–61. https://doi.org/10.14742/ajet.10050