IMPROVING ADULT EFL LEARNERS' SPEAKING PERFORMANCE THROUGH AI-ASSISTED ROLE-PLAY IN ONLINE LEARNING

Hoang Thi Hong Trinh
Department of Linguistics, Dong Nai Technology University, Dong Nai, Vietnam
Email: trinhhoang091098@gmail.com

Abstract: *This study examines the impact of AI-assisted role-play using ChatGPT on the speaking performance of adult EFL learners in an online environment. Using a mixed-methods quasi-experimental design, 28 learners were divided into an experimental group and a control group over eight weeks. Results show that the experimental group achieved significantly greater improvements in fluency and coherence. Qualitative findings indicate reduced speaking anxiety and positive perceptions of the AI tool's usefulness and ease of use, despite some limitations. The study suggests that AI-assisted role-play is an effective supplementary approach for enhancing speaking competence in adult EFL contexts.*

Keywords: *AI-Assisted Role-Play, speaking performance, adult EFL learners, Technology Acceptance Model 2, Perceived Usefulness, Perceived Ease of Use*

Nhận bài: 11/06/2026 *Phản biện: 11/07/2026* *Duyệt đăng: 14/07/2026*

I. INTRODUCTION

English plays a crucial role in global communication, yet many adult EFL learners continue to struggle with speaking fluency and coherence despite years of study . These difficulties are often linked to low confidence, speaking anxiety, and limited opportunities for authentic interaction, especially in online learning environments where direct communication is reduced .

Recent advances in artificial intelligence (AI), particularly tools such as ChatGPT, offer new possibilities for enhancing speaking practice through interactive and flexible role-play activities . However, empirical research on the effectiveness of AI-assisted role-play for adult learners in online contexts remains limited .

To address this gap, the present study examines the impact of AI-assisted role-play on adult EFL learners' speaking performance, focusing on fluency and coherence, and explores learners' perceptions based on the Technology Acceptance Model 2 (TAM 2) .

II. LITERATURE REVIEW

Recent studies highlight the importance of interactive approaches in improving EFL speaking skills, particularly through role-play activities that enhance learners' fluency, confidence, and communication strategies. With the rapid development of artificial intelligence, tools such as ChatGPT have created new opportunities for language learning by providing immediate feedback, flexible interaction, and authentic practice environments. Prior research indicates that AI-assisted learning can reduce speaking anxiety and increase learner engagement; however, most studies focus on general contexts rather than adult learners in online settings.

To better understand learners' acceptance of such technologies, the Technology Acceptance Model 2 (TAM 2) has been widely used to examine perceived usefulness and ease of use. Despite its relevance, limited research has combined AI-assisted role-play with TAM 2 in EFL contexts. This gap provides the basis for the present study.

III. METHODOLOGY

3.1. Research Design

To achieve a comprehensive understanding of the intervention, this research adopted a rigorous mixed-methods quasi-experimental design (Creswell & Creswell, 2018; Mackey & Gass, 2016). This design expertly combines the empirical strength of quantitative pre-test/post-test analysis to measure direct performance gains, alongside the deep explanatory power of qualitative interviews to unearth the affective and cognitive mechanisms driving those gains. The quasi-experimental nature was necessitated by the logistical constraints of intact online classes, meaning random assignment at the individual level was not feasible.

3.2. Research Site and Participants

The study was carefully situated at a prominent private English language center in Vietnam. This institution specializes in delivering online English communication courses utilizing platforms such as Zoom and Google Meet, thus providing an optimal, ecologically valid infrastructure for the digital intervention.

The participant pool comprised 28 adult EFL learners who were actively enrolled in intermediate online communication courses during the academic year. Purposive sampling was

meticulously applied to select these individuals, ensuring a homogenous sample suitable for comparative analysis (Etikan et al., 2016). The final valid sample consisting of 28 individuals was non-randomly but evenly assigned into an Experimental Group (TN, n = 14) and a Control Group (DC, n = 14). Selection criteria mandated that all participants were aged 18 or older, shared a verified comparable intermediate English proficiency level, and formally committed to completing all stages of the research.

3.3. Research Instruments

Data collection relied on highly validated quantitative and qualitative instruments. Speaking performance was quantitatively measured via structured, scenario-based speaking pre-tests and post-tests. The performances were critically evaluated using an analytical speaking rubric specifically adapted from standardized criteria, rigorously targeting Fluency and Coherence, and scored on a precise 1–5 scale. To guarantee scoring consistency and eliminate subjective bias, strict inter-rater reliability protocols were implemented utilizing two independent examiners. Qualitatively, in-depth semi-structured group interviews were conducted post-intervention with six selected participants from the experimental group to extract profound insights regarding their communicative anxiety, PU, and PEOU.

3.4. Experimental Intervention and Procedure

The intervention spanned a meticulously planned eight-week timeframe. Initially, participants in both groups completed the standardized speaking pre-test to establish baseline metrics. During the subsequent treatment phase, the Experimental Group (TN) participated in weekly, structured AI-assisted role-play sessions. They were tasked to interact vocally with ChatGPT (using voice-to-text/text-to-voice modalities) across various pragmatic scenarios (e.g., negotiating plans, expressing complex opinions, resolving customer service disputes). The AI provided immediate conversational feedback, follow-up interrogatives, and linguistic scaffolding.

Conversely, the Control Group (DC) received conventional, teacher-led online speaking instruction. This involved traditional break-out room role-plays and standard curriculum delivery entirely devoid of AI integration. Following the culmination of the eight weeks, both groups undertook the identical speaking post-test. Subsequently, the qualitative interviews were conducted with the experimental cohort.

3.5. Data Analysis

Quantitative data were systematically processed using SPSS software, a standard in behavioral sciences (Field, 2018). Independent-samples t-tests were executed to evaluate baseline equivalence between the groups. Paired-samples t-tests were subsequently utilized to measure within-group instructional gains from pre-test to post-test. The magnitude of the intervention's impact was calculated using Cohen's d effect sizes. Qualitative interview transcripts underwent rigorous, multi-phase thematic analysis (Braun & Clarke, 2006). This involved familiarization, initial coding, searching for themes, reviewing, and finally defining the recurring patterns regarding learner engagement and TAM 2 constructs.

Ethical Considerations

Ethical integrity was maintained with the utmost strictness throughout the study. Prior to any data collection, all participants were fully informed regarding the study's explicit purpose, procedures, and the nature of the AI intervention. Written informed consent forms were systematically collected from all 28 individuals. To ensure total confidentiality and protect participant anonymity, all identifying personal information was securely encrypted, and qualitative interviewees were assigned distinct alphanumeric thesis codes (e.g., Participant 1 mapped to HV09, Participant 2 mapped to HV08, up to Participant 6 as HV16) directly corresponding to the original raw data files (L.txt, Nh.txt, etc.). Furthermore, participants were unequivocally guaranteed the ethical right to withdraw from the research at any juncture without facing any academic or financial penalty.

IV. RESULTS

4.1. Quantitative Findings

4.1.1. Baseline Equivalence (Pre-test, EG vs. CG)

Prior to initiating the intervention, independent-samples t-tests were meticulously conducted to verify that the Experimental Group (TN) and Control Group (DC) possessed statistically equivalent speaking abilities, eliminating baseline bias. As detailed in Table 1, the mean baseline scores for Fluency were exactly identical across both cohorts ($M = 2.93$, $SD = 0.62$ for EG; $SD = 0.68$ for CG), yielding a p-value of 1.000. Regarding Coherence, the Experimental Group ($M = 3.07$, $SD = 0.33$) and Control Group ($M = 3.00$, $SD = 0.68$) exhibited no statistically significant difference ($p = .727$). These results robustly confirm there was no pre-existing advantage for either group before the AI intervention commenced.

Table 1. Baseline Equivalence (Pre-test, EG vs. CG)

Variable	EG Mean (TN)	CG Mean (DC)	Sig. (2-tailed)
Fluency_Pre	2.93	2.93	1.000
Coherence_Pre	3.07	3.00	.727

Pre-test vs. Post-test Scores by Group

Following the conclusion of the eight-week intervention, descriptive statistical analysis revealed observable improvements across both pedagogical settings. However, the raw gains observed in the Experimental Group were phenomenally superior. For the Fluency metric, the TN group dramatically improved their mean score by +1.43 points (from 2.93 to 4.36), in stark contrast to the DC group's moderate gain of +0.57 (from 2.93 to 3.50). Similarly, for the Coherence metric, the TN group advanced by an impressive +1.39 points (from 3.07 to 4.46), whereas the DC group progressed by +0.82 points (from 3.00 to 3.82). The detailed score trajectories are presented in Table 2.

Table 2. Pre-test vs. Post-test Scores by Group

Group	Measure	Pre-test M (SD)	Post-test M (SD)	Gain
TN (Experimental)	Fluency	2.93 (0.62)	4.36 (0.66)	+1.43
TN (Experimental)	Coherence	3.07 (0.33)	4.46 (0.46)	+1.39
DC (Control)	Fluency	2.93 (0.68)	3.50 (0.76)	+0.57
DC (Control)	Coherence	3.00 (0.68)	3.82 (0.61)	+0.82

Within-Group Gains and Effect Sizes

To ascertain the statistical significance of these internal improvements, paired-samples t-tests were executed. The results confirmed that both groups improved significantly on both speaking measures ($p < .05$). However, the most critical finding resides in the analysis of effect sizes (Cohen's d). The AI intervention's impact was remarkably profound. The TN group exhibited a mathematically "very large" effect on Fluency ($t = -8.27$, $df = 13$, $p < .001$, $d = 2.21$) and an "extremely large" effect on Coherence ($t = -14.91$, $df = 13$, $p < .001$, $d = 3.98$). These figures far surpass the conventional academic threshold for a large effect ($d = 0.8$), indicating a transformative pedagogical impact. The control group, while improving, showed notably smaller effect sizes (Fluency $d = 0.85$; Coherence $d = 1.77$).

Table 3. Within-Group Gains (Paired-Samples t-tests)

Measure	Group	t	df	p	Cohen's d
Fluency	TN	-8.27	13	<.001	2.21 (very large)
Fluency	DC	-3.17	13	.007	0.85 (large)
Coherence	TN	-14.91	13	<.001	3.98 (extremely large)
Coherence	DC	-6.62	13	<.001	1.77 (very large)

Post-test Comparison (EG vs. CG)

Directly addressing the primary research hypothesis (H1), an independent-samples t-test applied to the post-test scores demonstrated that the experimental group significantly and definitively outperformed the control group on both final metrics. The TN group scored significantly higher on Fluency (Mean Difference = +0.857, $p = .004$) and Coherence (Mean Difference = +0.643, $p = .004$).

Qualitative Findings

Rigorous thematic analysis of the six deeply semi-structured interviews (Participants P1 through P6) yielded three highly distinct overarching themes regarding learners' psychological experiences, cognitive load, and technological perceptions based on the TAM 2 framework.

Theme 1: Reduced Speaking Anxiety (Affective Filter)

The most prominent, uniformly and independently cited qualitative benefit was the profound reduction of the affective filter. Participants explicitly emphasized that interacting with ChatGPT entirely eliminated the paralyzing social pressure and fear of peer or teacher judgment

typical in human-to-human EFL practice. This psychological safety net drastically increased their Perceived Usefulness (PU). Furthermore, the 24/7 autonomous availability allowed for intensive, private rehearsal spaces. As Participant 1 (HV09) articulately expressed:

"ChatGPT is always ready to help, and I can practice anytime without feeling pressured, embarrassed, or afraid of making mistakes."

Theme 2: Usefulness Tempered by Limited Naturalness

While all participants universally acknowledged the AI's exceptional utility in expanding lexical resource, meticulously correcting grammar, and adopting highly varied personas (thereby cementing high PU scores), every single participant (6 of 6) identified stark limitations related to the naturalness of the generated language. They astutely noted the AI's structural tendency to sound overly academic, formulaic, or insufficiently personalized for casual discourse. Furthermore, tangible issues with Perceived Ease of Use (PEOU) surfaced prominently when learners attempted to handle highly complex, multi-turn prompts.

"It's fun because ChatGPT can play many different roles-for example, when I asked it to act like a middle-aged person using old-fashioned language, it gave me really interesting responses." (Participant 6, HV16)

"When I ask more advanced questions, ChatGPT gets confused and steers me toward directions I didn't intend." (Participant 5, HV15)

Theme 3: A Preference for Blended Learning

Critically, despite the massive quantitative gains, no participant advocated for the complete pedagogical replacement of human educators by AI. The unyielding consensus was that generative AI serves as an exceptional, low-stakes supplementary tool for extensive rehearsal and confidence building. However, human teachers were deemed absolutely irreplaceable for essential phonetic correction (e.g., physically observing mouth shapes for articulation), conveying authentic emotional nuance, and providing the unpredictable, genuine conversational feedback required for true fluency.

V. DISCUSSION

The meticulous integration of quantitative and qualitative data provides a highly robust, triangulated answer to both primary research questions, offering significant novel contributions to the field of AI-assisted language learning.

5.1. Discussion on RQ1: Speaking Performance Gains

Regarding RQ1, the empirical data conclusively demonstrate that AI-assisted role-play dramatically and significantly improves adult learners' speaking fluency and coherence compared to traditional online instruction. These findings strongly corroborate, yet exponentially exceed, the earlier foundational studies on standard role-play efficacy established by Aliakbari and Jamalvandi (2010), Cho (2015), and Krebt (2017).

A highly novel and critical finding of this specific study is the observed discrepancy in effect sizes between the two speaking metrics. The quantitative analysis revealed a substantially higher effect size in Coherence ($d = 3.98$) than Fluency ($d = 2.21$). This statistical anomaly perfectly mirrors the qualitative interview data, wherein four out of six participants specifically credited the text-based logic mapping of ChatGPT with helping them visually and cognitively organize their ideas, arguments, and vocabulary schemas-the very core definition of coherence. Because ChatGPT processes and outputs highly structured, logical text, learners inadvertently absorbed this structural integrity, improving their logical coherence at a faster rate than their raw, physiological speaking speed.

These findings align seamlessly with Cognitive Learning Theory. By offering immediate, impeccably structured feedback and effortlessly adopting various conversational personas, ChatGPT functions flawlessly as an effective, tireless scaffold within the adult learner's ZPD (Vygotsky, 1978). Concurrently, the AI's capability for limitless, low-stakes, anxiety-free repetition fundamentally supports behavioral habit formation (Skinner, 1957) and facilitates constant exposure to comprehensible input finely tuned to the user's level (Krashen, 1985).

5.2. Discussion on RQ2: Learner Perceptions and TAM 2

Regarding RQ2, analyzed deeply through the lens of TAM 2 (Davis, 1989; Venkatesh & Davis, 2000), the intervention successfully generated extraordinarily high levels of Perceived Usefulness (PU). The qualitative data isolated the primary mechanism driving this PU: the dramatic reduction in speaking anxiety. Foreign

language anxiety is a well-documented, severe barrier in adult EFL acquisition (Horwitz et al., 1986; Woodrow, 2006). Because the learners felt completely unjudged by the machine, their affective filter was decisively lowered, prompting them to willingly practice significantly more frequently and with drastically less avoidance behavior than they would exhibit in a traditional online classroom setting (Shen et al., 2025). This massive increase in practice volume directly lowered cognitive load during speech production (Sweller, 1988), explaining the sheer magnitude of the quantitative gains.

However, the study also reveals crucial nuances regarding Perceived Ease of Use (PEOU). While technically accessible, the AI's PEOU was practically hindered by its algorithmic constraints. The qualitative critique regarding the AI's "formulaic language" and occasional cognitive breakdown during complex multi-turn prompts serves as a vital caveat. It suggests that the unusually large quantitative gains ($d > 3.0$) might partially reflect the learners' improved performance on highly structured, rubric-aligned academic tasks, rather than a mastery of fully naturalistic, unpredictable, spontaneous human fluency. This critical insight underscores the learners' qualitative conclusion: AI is a phenomenally powerful rehearsal space, but its linguistic outputs are not yet a substitute for authentic human interaction.

5.3. Novel Contributions of the Study

This research provides several novel contributions to the literature. First, it quantitatively proves that generative AI has a disproportionately higher impact on adult learners' coherence compared to their fluency, challenging the assumption that chatbots primarily boost speaking speed. Second, it theoretically extends the application of TAM 2 in educational technology by demonstrating that, for adult language learners, the reduction of the affective filter (anxiety) is the single most dominant variable influencing Perceived Usefulness. Finally, it provides empirical backing for a specific pedagogical model: AI as a primary rehearsal tool, and human teachers as the ultimate arbiters of phonetic and emotional accuracy.

5.4. Limitations and Future Research

Despite these profound benefits and strong effect sizes, the study acknowledges certain limitations. The sample size of 28 participants, while statistically valid for the methodologies employed, restricts broad global generalizability. Additionally, the qualitative feedback highlighting the occasionally unnatural, overly academic tone of current AI responses explicitly indicates that AI cannot yet fully replicate authentic human interaction. Future research should aim to conduct longitudinal studies with larger, more diverse adult cohorts across different proficiency levels. Furthermore, as LLM technology rapidly evolves, future studies should specifically investigate whether newer, voice-native multimodal AI models can successfully overcome the current limitations regarding naturalness and emotional phonetic feedback.

VI. CONCLUSION

This comprehensive mixed-methods study robustly concludes that intelligently integrating ChatGPT-assisted role-play into online learning environments yields statistically significant, practically transformative improvements in both the fluency and coherence of adult EFL learners. These profound improvements are primarily mediated by the technology's unique ability to provide a zero-pressure, highly accessible, and infinitely patient interactive environment that dramatically lowers speaking anxiety, thereby maximizing voluntary practice opportunities and cognitive engagement.

The findings carry significant theoretical and practical implications. Theoretically, the study reinforces the validity of combining Cognitive Learning paradigms with TAM 2 to understand AI adoption in adult education. Practically, it is highly recommended that language educators and curriculum designers proactively adopt a blended, synergistic pedagogical approach. AI-assisted role-play should not be viewed as a threat, but rather systematically integrated into syllabi as a mandatory, low-stakes supplementary tool specifically targeted for vocabulary building, structural coherence generation, and initial confidence building prior to live classes. Concurrently, human instructors must remain the central, irreplaceable pillar of the curriculum, pivoting their focus away from basic vocabulary drills and towards guiding complex phonetic pronunciation, emotional prosody, and navigating the nuances of genuine, unpredictable human conversation.

REFERENCES

Aliakbari, M., & Jamalvandi, B. (2010). The impact of role play on fostering EFL learners' speaking ability: A task-based approach. Journal of Pan-Pacific Association of Applied Linguistics, 14(1), 15–29.

British Council. (2013). The English effect: The impact of English-What it's worth to the UK and why it matters to the world. https://www.britishcouncil.org/research-insight/english-effect

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). Sage.

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Krebt, D. M. (2017). The effectiveness of role play techniques in teaching speaking for EFL college students. Journal of Language Teaching and Research, 8(5), 863–870. http://dx.doi.org/10.17507/jltr.0805.04

Mohammed, G. M. S., & Mohammed, S. (2021). Speaking skills in online learning: An investigation of the strategies used by EFL learners at the University of Bisha. Asian EFL Journal, 28(2.3), 120–134.