ISSUES IN LANGUAGE TEACHING
2026, VOL. 15, NO. 1, 213–241
https://doi.org/10.22054/ilt.2026.92507.987

ALLAMEH TABATABA'I UNIVERSITY PRESS

The Effects of AI-Powered Chatbots vs. Synchronous CMC on Oral Proficiency and Self-Efficacy in Under-Resourced TESOL

Mohammad Mahdi Hajmalek [1] and Negin Etedal [2]

1. *Corresponding Author,* Department of English Language, Faculty of Humanities, Khatam University, Tehran, Iran. E-mail: m.hajmalek@khatam.ac.ir
2. Department of English Language, Khatam University, Tehran, Iran. E-mail: n.etedal@khatam.ac.ir

ABSTRACT

Regardless of the current buzz, there is still a great deal to uncover when it comes to the promises of artificial intelligence (AI) in language instruction, especially in under-resourced contexts. The present mixed-methods study compared the effects of AI-powered chatbots and Synchronous Computer-Mediated Communication (SCMC) on Iranian English as a foreign language (EFL) learners' oral proficiency and self-efficacy. For this purpose, 60 EFL learners were conveniently sampled and the speaking section of the international English language testing system (IELTS) as well as the Questionnaire of English Self-Efficacy were administered as pretest and posttest with six weeks of treatment. The experimental group used Pi as speaking partner, whereas the other group spent a comparable amount of time engaging in SCMC with the instructor. Analysis of covariance (ANCOVA) revealed that AI significantly enhanced oral self-efficacy in the experimental group ($p = 0.001$). As for oral proficiency, although a positive impact was suggested, the difference was only marginally significant ($p = 0.082$). Qualitative findings mostly suggested positive attitudes, particularly toward the "increased practice opportunities" while a number of concerns, especially about "limitations in natural interaction," were voiced. It is argued that AI has the potential to compensate for some of the limitations in under-resourced teaching English to speakers of other languages (TESOL) contexts and push language education to more equitable and sustainable standards around the globe.

ARTICLE TYPE:
Research Article

ARTICLE HISTORY:
Received 02 March 2026
Received in revised form 03 June 2026
Accepted 15 June 2026
Published Online 30 June 2026

KEYWORDS:
Artificial intelligence, conversational chatbots, EFL learners, oral proficiency, self-efficacy

CITE THIS ARTICLE: Hajmalek, M. M., & Etedal, N. (2026). The Effects of AI-Powered Chatbots vs. Synchronous CMC on Oral Proficiency and Self-Efficacy in Under-Resourced TESOL. *Issues in Language Teaching*, *15*(1), 213–241. https://doi.org/10.22054/ilt.2026.92507.987

Introduction

Artificial intelligence (AI) is already securing a solid position in language teaching and learning due to the several advantages it offers. Language educators are, hence, under growing pressure to integrate AI into learning experiences and pave the way for the benefits of a "synergized" model of in-class and out-of-class learning (Lai & Sundqvist, 2025). Building upon the achievements in Large language models (LLMs), Natural language processing (NLP), and deep learning, chatbots have recently attracted enormous attention as convenient and stress-free pedagogical tools (Kang & Sung, 2024; Kim et al., 2022). AI has been demonstrated to contribute to a variety of second language (L2) skills and elevate students' learning experience by enhancing their affective variables, including motivation, self-efficacy, anxiety, and self-regulated learning (Huang & Mizumoto, 2025; Liu & Reinders, 2025; Yamaoka, 2024). Furthermore, within the interactionist (Long, 1996) and sociocultural (Lantolf, 2000) frameworks of language learning, AI interactions can be hypothesized to compensate for limitations in accessible and sustainable exposure to input, scaffolding feedback, and opportunities for output (Du & Daniel, 2024; Ji et al., 2022).

Several studies have already acknowledged the possibility of using AI chatbots to develop the oral skills of English as a foreign language (EFL) learners (e.g., Huang, 2024). Chatbots seem to provide excellent opportunities for informal digital learning of English (IDLE), as emphasized in the literature (Fryer et al., 2020; Godwin-Jones, 2021). However, AI conversational partners might not be solely adequate in guaranteeing uninterrupted language learning beyond the classroom (Guan et al., 2025). This is especially true since earlier modes of technology-enhanced L2 learning practices, including Computer-Mediated Communication (CMC) and Synchronous Computer-Mediated Communication (SCMC), have already demonstrated great potential in the literature (e.g., Jiang & Chun, 2023; Ziegler, 2016). Yamaoka (2024), for example, suggested that ChatGPT could lead to reduced confidence if not employed cautiously. Despite all the promises of AI in language teaching, the integration of such technologies into educational systems, especially under-resourced ones, as well as their possible benefits and challenges, seem to call for more complex educational, social, contextual, and ethical considerations.

In line with such critiques, Todd (2025) recommended a critical attitude toward AI in language teaching and suggested that "being prepared allows the field to resist changes that could ultimately be detrimental" (p. 37). Tin (2025) also warned against the possible toll of AI on creativity in language

education. Liu and Reinders (2025) pointed at the relative scarcity of empirical research in the field, especially regarding "learners' perception". They suggest that "there is a clear need for more systematic investigations into specific areas of real-world challenges where AI can be hypothesized to offer an advantage over current practices" (Liu & Reinders, 2025, p. 1). Godwin-Jones et al. (2024) also called for a "critical, balanced approach to AI integration in L2 instruction", taking AI's both advantages and drawbacks into account (p. 9). Fryer et al. (2020) noticed that despite the existing literature, "scant progress toward chatbots as substantive language learning partners has been made" (p. 17). This observation is important not only in terms of the pedagogical outcomes but also with regard to the affective dimensions of modern L2 education. Therefore, several pedagogical, technological, and social variables warrant further, more in-depth exploration (Jeon, 2024).

Among such variables, self-efficacy has been often implicated as a major determining factor in language learning success (Teng et al., 2023). While self-efficacy beliefs predict learners' use of metacognitive strategies, such positive attitudes also lead to improved motivation and perceived progress in online language learning contexts (Teng, 2025). However, there have been some doubts about chatbots' ability in replicating the emotional engagement and sensitivity often found in human interactions (Wiboolyasarin et al., 2025). Therefore, exploring the detailed effects of AI chatbots on self-efficacy, especially when it comes to interactive and socially oriented skills such as speaking, offers worthwhile insights.

Finally, Lai and Sundqvist (2025) emphasized a more profound consideration of contextual factors, including a particular focus on "underprivileged regions and contexts" (p. 18) in AI research. The switch to technology-powered language learning, particularly in the shadow of the recent pandemic, presented a serious challenge not only to under-resourced educational systems but also to all contexts where stakeholders were long accustomed to traditional education, such as China (Teng, 2024), Iran (Ghiasvand et al., 2024), and many similar countries. Therefore, given the number of settings where language educators and learners are struggling with limited resources or heavily traditional practices, it is important to probe the blessings and curses of AI for ensuring affordable, sustainable, and equitable language learning.

Given this introduction, the present study sought to compare the affordances of AI chatbots as speaking partners with SCMC, as an established human-delivered interactive technology for language learning, in promoting EFL learners' oral proficiency and self-efficacy in an under-resourced

teaching English to speakers of other languages (TESOL) context. The paper will also present arguments regarding what we call under-resourced language education and what kinds of settings could be classified under that term, especially in a post-pandemic world, packed with technological options, yet, diverse in resource distributions. The following research questions were posed for this purpose:

1. Does the use of AI conversational chatbots make a significant difference in improving EFL learners' oral proficiency and self-efficacy as compared with SCMC?
2. How do EFL learners in an under-resourced TESOL context perceive the advantages and drawbacks of AI-powered chatbots as speaking partners?

Literature Review

AI Chatbots and L2 Oral Skills

AI has revolutionized or, to be more discreet, is revolutionizing language learning in virtually all aspects and L2 speaking is no exception. Koç and Savaş (2025) reported that the body of research between 2010 and 2024, even before the prevalence of AI, mostly confirms the capabilities of voice-based chatbots in developing L2 skills and improving learners' affective variables, including motivation, willingness to communicate (WTC), and anxiety. Similarly, Liu et al.'s (2026) meta-analysis of 25 papers published on the topic between 2022 and 2024 revealed that AI-powered chatbots were generally effective in promoting L2 learning, positively influencing learners' academic performance and psychological well-being.

Kim et al. (2022) modelled a conversational chatbot design which led to significant improvements in EFL learners' experience of language learning, including a sustained length of conversation, engagement in problem-solving, and substantiated meaning negotiation. Shin and Lee (2024) compared the effectiveness of three chatbots as speaking partners, including ChatGPT, Poe AI, and *Pi*. Kang and Sung (2024) employed an AI chatbot to track changes in EFL learners' conversational skills and found it significantly efficient in improving learners' speaking as well as self-directed learning. Peng and Liang (2025) showed that although the link between learners' WTC and use of chatbots can be dynamic and complex, participants generally showed positive attitudes and increased WTC when interacting with ChatGPT. Wu et al. (2025) found the integration of AI chatbots beneficial in decreasing L2 learners' speaking anxiety and improving enjoyment.

Particularly, in the Iranian EFL education, as the context of the present study, Shafiee Rad (2024), who employed Speeko as an AI-powered speaking app, found positive results on learners' speaking proficiency, willingness to communicate, and attitudes. Similarly, Baharloo and Miyan Baghi (2024) confirmed the benefits of Gliglish and Sayra apps on Iranian EFL learners' speaking proficiency and found very positive attitudes on their side. Gholami et al. (2025) concluded that learners with higher levels of digital literacy would benefit more from Anima chatbot in improving their oral skills.

However, despite the general optimism toward chatbots as learning partners, especially in earlier studies, it seems that more discreet critiques are emerging in the field. For instance, although systematic reviews such as Koç and Savaş (2025), Liu et al. (2026), Lyu et al. (2025) and Yang and Li (2024) have unanimously found generally positive effects for chatbots in L2 learning, they have also noticed a number of caveats and concerning patterns. Koç and Savaş realized that some learners had complained about how interactions with chatbots felt unnatural and mechanical, and many users' motivation to use AI wore off as the initial enthusiasm faded away. Liu et al. (2026) also raised concerns regarding the rigor of some unconditional claims to the efficiency of AI tools and called for more cautious interpretation of the results. Lyu et al. (2025), who specifically focused on the effectiveness of chatbots in their meta-analysis of 31 studies, concluded that in order to unleash the full potential of AI chatbots, their accessibility, modality, and adaptability must be taken into consideration. This conclusion is of very relevant importance in the integration of AI chatbots in under-resourced TESOL due to its emphasis on contextual considerations.

Reviewing 100 peer-reviewed papers between 2009 and 2024, Du and Reynolds (2025) point to the general usefulness of chatbots, but simultaneously call for a more critical integration of AI into language learning programs. They mention hallucinations, bias, and ethical risks among topics less often attended to in chatbot research. Joo (2024) also emphasizes contextual factors along with proficiency, elicitation (task type), socio-cognitive, instructional, affective, social-interactional, and technological dimensions in the efficacy of AI as writing and speaking partners. This eight-dimensional model also highlights the role of learners' affective and social variables, including their sense of self-efficacy.

Computer-Mediated Communication (CMC) and AI

CMC, broadly defined as any human-to-human interaction on a digital platform, has attracted language educators' attention since the introduction of

Web 2.0 technology (Ziegler, 2016). In fact, pioneering studies highlighting the applicability of CMC in L2 learning started to emerge as early as the 2000s (e.g., Abrams, 2003; Smith, 2003). Smith (2003) demonstrated how CMC can fit into Long's (1985) interactional hypothesis within a technologically updated modality. CMC features both synchronous and asynchronous options, referring to live chats (real time) and offline communications (delayed) respectively. Jiang and Chun (2023) showed the superiority of CMC in developing learners' speech comprehensibility and confidence in comparison with traditional methods.

While the benefits of CMC for L2 oral proficiency development were already established in the pre-AI era (e.g., see Lin, 2015), the introduction of AI raises the question whether human-machine interactions can be actually as useful as technology-mediated human-human interaction. The public debut of chatbots powered by LLMs has opened a new realm of unprecedentedly intelligent communication *with* technology rather than *through* technology (Hancock et al., 2020). Therefore, it might seem timely to switch from discussions of SCMC to technology-led interactions and AI-controlled speaking practice, focusing on the affordances of this new technology.

The Significance of Self-efficacy for Speaking Development

Technology-powered language teaching calls for updating our perspectives on learners' and educators' emotions and attitudes. Self-efficacy (SE) was famously popularized in Bandura's (1977) work on social cognitive theories and soon found its way into educational and language learning discussions due to its role in predicting individuals' success in fulfilling tasks. Learners with higher degrees of SE are more likely to engage in a task, persist in the face of challenges and, generally, invest more effort (Leeming et al., 2024). Several factors can affect language learners' SE. Particularly in digital education, Liu and Zhang (2025) showed that learners with higher levels of New Media Literacy enjoyed elevated self-efficacy. If a direct link between motivation and SE is assumed, Liu and Reinders' (2025) study showed that L2 learners' interaction with an AI-powered chatbot could significantly boost their motivation as compared with a pre-scripted chatbot. Leeming et al. (2024) showed that speaking self-efficacy, along with perceived communicative competence, and WTC predicts L2 oral production. Therefore, a more solid base of research on the role of AI in learners' self-efficacy in speaking is called for.

Under-resourced Language Teaching and Learning

Bailey and Christian (2021) justifiably argue that "recommendations from research and experience in well-resourced contexts may not necessarily transfer in a straightforward manner to other situations" (p. 1) and call for highly contextual studies that promote creative and innovative approaches and draw on local assets to set effective policies and practice standards. When discussing the applicability of such technological advancements as AI in under-resourced contexts, revisiting established frameworks in the field, such as TESOL technology standards (Healey et al., 2011) seems indispensable. Introduced in 2008, the standards offered a definition of low-resource, mid-resource, and high-resource settings. Low-resource contexts, back then, were defined as those with limited access to computers and internet. After nearly two technologically packed decades and a game-changing global pandemic, it can be argued that revisions to those definitions are due (see Hubbard (2021) as well as Karatay and Hegelheimer (2021) for arguments and counterarguments).

More recently, in an effort to define under-resourced contexts in language learning, Curtis (2021) resorted to standards from three international organizations, namely, The Organization for Economic Co-operation and Development, World Bank, and the United Nations, and concluded that, while the concept can be dynamic and slippery, a hierarchical framework can be set to classify countries in terms of their resources. Building on those grounds, the present study argues that TESOL contexts such as Iran, despite the abundance of computers and smartphones and access to the internet, can be considered as technologically under-resourced language teaching settings, if not low-resource ones. In such contexts, the resources might not be *low* per se, but are arguably lower than the requirements of a smooth L2 learning experience in the third decade of the 21st century. Although there is internet access, full and legitimate connectivity to global app markets is occasionally limited due to both national and international restrictions, such as the limitations in monetary transactions. AI chatbots, particularly, are only marginally accessible and there are barriers in purchasing full or unlimited versions.

Hedayati and Marandi (2014) argued that categorically there were three barriers against the integration of computer-assisted language learning (CALL) in Iranian EFL context: *facility*, *teacher*, and *learner* constraints; this can be argued to apply to the current status of AI in Iranian English language teaching (ELT) context after more than a decade. They listed poor internet connectivity, inadequate quantity or quality of technological tools, service

restrictions, teachers' lack of training and/or inhibitive attitudes, lack of support from administrators, lack of digital literacy among instructors and learners, and learners' lack of autonomous learning behavior among the reasons why CALL integration was not fully realized in the Iranian context. More recently, Moslemi Nezhad Arani et al. (2024) made very similar observations regarding the status of CALL in Iran and mentioned institutional support policy and learner variables such as level of digital literacy, cultural and religious background, learning styles, and age, as obstacles against the successful implementation of CALL in Iran. Ghiasvand et al. (2024) also concluded that Iranian EFL teachers do not feel "ready" for fully accepting AI-mediated instruction due to its economic requirements and technological infrastructures.

Furthermore, in the Iranian ELT context, global connectivity and use of English as a Lingua Franca for commerce, education, industry, and entertainment on a daily basis are limited or in many cases non-existent. Although technological options are available, their normalized integration into the educational system in general, and language teaching in particular, is still in its infancy and many instances can be categorized as what is known as behavioristic CALL rather than the integrative one. Furthermore, in many settings, traditional education is still preferable by older teachers, who are mostly digital immigrants, and administrators, whose hands are tied when it comes to the costs of renovating resources and practices.

One solution to dealing with such limitations in under-resourced contexts, can be offered by AI technologies. Guan et al. (2025) assert that the integration of AI in language education can offer not only pedagogical and technological advantages but also humanistic ones. They argue that "Humanistically, such technology could advance educational equity by preventing student characteristics from negatively interacting with and influencing the learning resources and environments" (Guan et al., 2025, p. 327). Furthermore, Suryanarayana et al. (2024) argue that as the use of AI in education can improve "student results, learning experiences, and administrative processes", it is an ideal candidate for ensuring sustainability in education. Lin et al. (2023) propose that AI obviously offers the potential to address the barriers sustainable education is facing, including "inadequate infrastructure, limited resources, and a lack of awareness and engagement" (p. 1). The newly emerged AI technologies can serve the goal of educational and financial empowerment in underdeveloped and under-resourced educational systems around the globe (Shabbir et al., 2024) and help build sustainable, equitable, accessible, and fair language education.

Method

Participants

For this explanatory sequential mixed-methods study, 60 EFL learners, (38 female and 22 male), with an average age of 26, from a pool of learners in a language institute in Iran were conveniently sampled. The Quick Oxford Placement Test (QOPT) was administered to the available cohort to ensure that outliers were excluded and all participants demonstrated reasonable L2 proficiency to engage in independent conversational chats with AI. By selecting learners falling within ± 1 standard deviation (SD) from the mean, 60 learners were randomly and equally assigned to the Chatbot and SCMC groups. For the qualitative phase of the study, 10 volunteers from the Chatbot group were interviewed. In order to obtain informed consent, the participants were briefed on the general purpose of the study and the fact that they could benefit from extracurricular practice if they agreed to participate in the experiment.

Instrumentation

Other than the QOPT, used to measure baseline proficiency levels, the following instruments were used for data collection.

Speaking pretest and posttest

The speaking section of IELTS and its public rating rubrics were used. The prompts were adapted from sample tasks in the Cambridge IELTS series and Expert opinions were elicited to ensure the suitability of the task level and topics for learners in this context. Insights from the pilot phase were used to fine-tune the selections. For the second task of the test, initially 10 different topics were extracted from the samples with a collection of possible follow-up questions for the third task. After the briefing session with two experienced IELTS instructors, two of the topics were finalized for the oral pre- and posttests.

Self-efficacy

Wang et al.'s (2013) Questionnaire of English Self-efficacy (QESE), was adopted to assess participants' self-perceived abilities in English. QESE consists of 32 items on a 7-point Likert scale and asks respondents to judge their capabilities in fulfilling a variety of tasks in English. Two experts were consulted to ensure the content validity of the questionnaire and its applicability to the context of the study. Cronbach's alpha values calculated

for the overall scale (α = .94) and the subscales indicated a high level of internal consistency.

Interviews

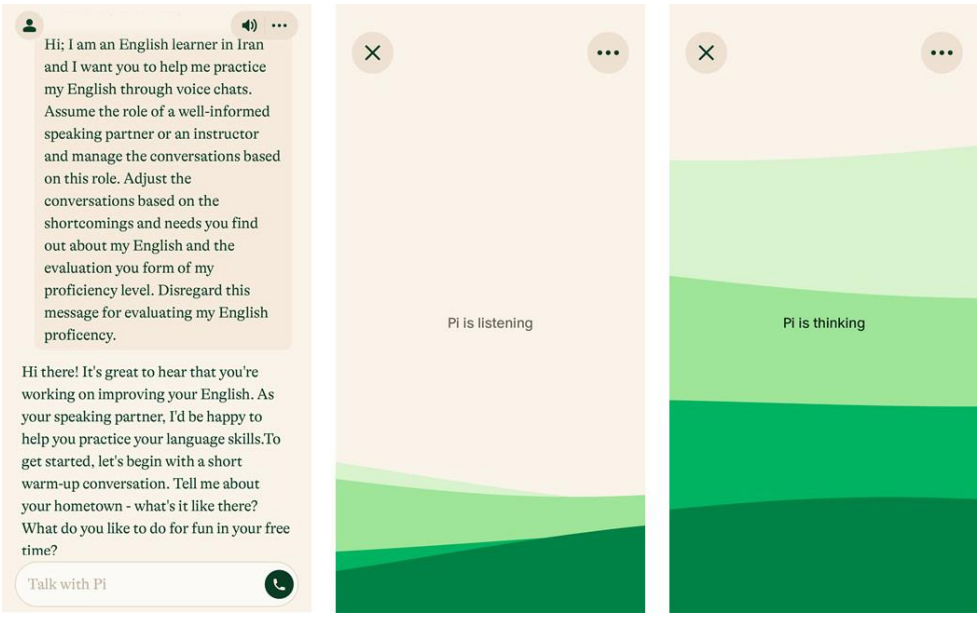
For the qualitative phase of the study, a semi-structured interview was designed to explore learners' experience with the chatbot. For this purpose, 10 baseline questions were developed after exploring the literature, to elicit learners' perception of the uses, usefulness, and challenges of the approach (See Appendix A for the English translation). The interview was finalized after consulting a panel of three experts about the organization, content, and depth of the questions and minor modifications were made. In order to enhance the trustworthiness of the findings, respondent validation was also performed after data collection by asking the participants to review a summary as well as an initial coding of their contributions and confirm or rectify the researchers' interpretations.

Data Collection Procedure

The entry oral proficiencies were measured through an online administration of the IELTS speaking test, screen-recorded with consent from the participants for rating purposes. Responses were then rated by two independent and experienced IELTS instructors, who previously had a briefing session to finalize the topics and form consensus over the use of rating rubrics. The inter-rater reliability checks were performed to ensure consistency between raters' scorings, which yielded intraclass correlation (ICC) values of .91 and .94 for pretest and posttests, respectively. The QESE was also administered to record the base-line self-efficacy scores, which showed a very high internal consistency of .96 based on Cronbach's Alpha.

For AI-assisted speaking practice, after a piloting phase, the *Pi* chatbot was employed for several reasons. Most importantly, the app was easily available for free, supported voice-based interactions, offered opportunities to engage in meaningful, real-time conversations on a variety of topics, and provided a user-friendly and smooth user experience with a simple interface. Additionally, *Pi* is capable of transcribing the conversations and adopting real-time feedback mechanisms that pinpoint errors in grammar, pronunciation, and word choice, while providing suggestions for improvement, fostering self-monitoring and autonomous learning. Figure 1 reproduces screenshots from *Pi* application.

Figure 1:
Sample Screenshots of *Pi* Chatbot User Interface



During the treatment, the chatbot group regularly used *Pi* out of the classroom after clear guidelines were delivered to them. The instructor also assisted them with downloading and running the app on their smartphones and demonstrated its use. A fixed prompt was provided by the instructor (Figure 1), to get the chats started, but the learners were given guidelines on how to maintain the conversations in future. The playback speed was set at 1x (normal) for all learners and they were helped to select the voice and accent of their choice. Participants in this group were required to converse with *Pi* for at least 15 minutes per session. The application provided a range of discussion topics, but they were also given a list of topics similar to those used in the SCMC group for reference when they felt they wanted more control. The instructor regularly checked learners' activity by following up on their exercises every session and checking their chat history while providing technical tips whenever necessary. However, the learners were assured that their chats would remain private and the instructor did not read the details of the transcripts.

The SCMC group interacted with the instructor via Google Meet for approximately 15 minutes per session. The teacher facilitated the conversations, asked questions, and provided feedback on participants' language use in real time. After 12 sessions of treatment, posttests were

administered following a similar procedure as the pretest. The six weeks of treatment were hoped to reduce the chances of history effect for the re-administration of QESE. Furthermore, online interviews were held with volunteers from the Chatbot group and the iterative process of data collection, data screening, and initial coding continued until data saturation was ensured with the tenth participant. The interviews were held in users' native language (Persian) and each session lasted for 20 minutes on average, based on the respondents' willingness to share details. To further assure the trustworthiness of the qualitative phase, member checking, as a form of respondent validation, was conducted by providing the participants with a summary of their insights and asking them to confirm the content. The qualitative data was coded by two coders after ensuring inter-coder consistency and discussion sessions were held to exchange insights and achieve final consensus on the organization of the themes. Intra-coder reliability checks were also performed to ensure each coder's consistency over time.

Data Analysis

In the quantitative phase, two sets of analysis of covariance (ANCOVA), with a follow-up Bonferroni-adjusted pairwise comparison were employed to compare the effects of the treatment type on learners' oral proficiency and self-efficacy, while controlling for covariates from the pre-intervention stage. For the qualitative research question, inductive thematic analysis was employed in order to avoid biases and make sure that the qualitative findings are based on participants' actual experiences in the present study.

Results

Quantitative Findings

First, the normality of the data and the equality of variances, as assumptions of ANCOVA, as well as the reliability of the scales and inter-rater consistency (as reported earlier) were ensured. Descriptive statistics of the dependent variables in posttests are displayed in Table 1 below.

Table 1.
Descriptive Statistics of Posttest Oral Proficiency and Self-efficacy

Measure	Grouping	Mean	Std. Deviation	N
Oral Proficiency	Chatbot	4.67	1.44	30
	SCMC	3.89	1.26	30
	Total	4.28	1.40	60
QESE	Chatbot	6.03	0.76	30
	SCMC	4.93	1.54	30
	Total	5.48	1.32	60

Table 1 reveals a slight advantage in oral proficiency for the Chatbot group with a marginally higher variability within the group. However, inferential statistics were needed to check the significance of these differences. Table 2 reports the results of the one-way ANCOVA for oral proficiency, while controlling for baseline proficiency as measured through QOPT. Given the randomized assignment of learners with differing pretest oral proficiency ratings to the experimental groups, entry level QOPT scores were taken into consideration as a covariate to control for the overshadowing effects of overall proficiency, which was hypothesized to be a more determining factor in learners' multilevel interaction with AI. Furthermore, this allowed the analysis to avoid overcorrecting for performance on a task closely aligned with the posttest measure.

Table 2.
Tests of Between-Subjects Effects for Oral Proficiency

Source	Type III Sum of Squares	df	Mean Square	F	*p*
Corrected Model	14.869[a]	2	7.434	4.188	0.020
Intercept	127.796	1	127.796	71.993	0.000
Grouping (AI Tools)	5.553	1	5.553	3.128	0.082
Baseline Proficiency	10.916	1	10.916	6.149	0.016
Error	101.182	57	1.775		
Total	1217.480	60			
Corrected Total	116.051	59			

a. R Squared = .128 (Adjusted R Squared = .098)

The analysis revealed that the overall corrected model, which includes the use of AI tools (Group) and baseline proficiency, significantly predicted speaking scores (p = .020), suggesting that approximately 13% of the variance in the posttest was explained by the model. However, although the p-value of .082 for the Group alone suggests a trend and hints at a possible advantage, it does not quite reach conventional levels of significance. All in all, the baseline

proficiency proved to be a more significant covariate in determining learners' post-intervention performance on the oral tests, while the marginal group effect shows a promising role for AI speaking partners in a larger sample or more extended treatment periods.

As for self-efficacy, it is evident in Table 1 that the Chatbot group (M = 6.03) ended up with a higher mean than the SCMC (M = 4.93), with a lower SD. To run the inferential statistics, first, the assumption of homogeneity of variances was confirmed through Levene's test and then a one-way ANCOVA was conducted to compare groups' self-efficacy scores on the posttest after adjusting for pretest SE levels. Table 3 displays the results of ANCOVA.

Table 3.

Tests of Between-Subjects Effects for Self-efficacy

Source	Type III Sum of Squares	df	Mean Square	F	*p*	Partial Eta Squared
Corrected Model	48.07	2	24.03	24.50	0.000	0.46
Intercept	39.01	1	39.01	39.77	0.000	0.41
Baseline SE	29.92	1	29.92	30.50	0.000	0.34
Group	12.96	1	12.96	13.22	0.001	0.18
Error	55.90	57	0.98			
Total	1908.00	60				
Corrected Total	103.98	59				

R Squared = 0.46 (Adjusted R Squared = 0.44)

As shown in Table 3, the corrected overall model is strongly significant, $F(2, 57) = 24.50$, $p = 0.000$, explaining approximately 46% of the changes in post-intervention SE levels, which reflects a large effect size. It can be inferred that baseline SE scores significantly predicted posttest SE, $F(1, 57) = 30.50$, $p < .001$, with a large effect size ($\eta_p^2 = .34$), and the grouping variable also demonstrated a significant effect, $F(1, 57) = 13.22$, $p = .001$, with a moderate effect size ($\eta_p^2 = .18$). This indicates that while learners with higher levels of initial self-efficacy tended to maintain or enhance their SE throughout the intervention, the participants who interacted with the AI chatbot reported significantly higher self-efficacy levels post-intervention, even after adjusting for baseline SE scores. Table 4 presents the results of Bonferroni-adjusted pairwise comparisons.

Table 4.
Pairwise Comparisons for Self-efficacy

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	*p*	95% Confidence Interval	
					Lower Bound	Upper Bound
Chatbot	SCMC	0.936*	0.257	0.001	0.421	1.452
SCMC	Chatbot	-0.936*	0.257	0.001	-1.452	-0.421

* The mean difference is significant at the .05 level.
Adjustment for multiple comparisons: Bonferroni.

In sum, the results of post-hoc comparisons confirm that the use of AI chatbots significantly enhanced the oral self-efficacy of the participants.

Qualitative Findings

In the qualitative phase, the present study explored learners' perceptions of the advantages and drawbacks of using an AI-powered chatbot as a speaking partner, regarding their oral proficiency and self-efficacy. After thematic analysis of the transcribed interviews, the perceived benefits were postulated in four main themes with a total of 10 subthemes and drawbacks were organized into five major and 12 minor themes, as summarized in Tables 5 and 6 below, respectively.

Table 5.
Perceived Benefits of AI-Powered Chatbots for Oral Language Development

Theme	Subtheme	f	F (total)
Increased Practice Opportunities	ubiquitous access	9	22
	increased practice time	8	
	personalized, adaptive practice	5	
Reduced Anxiety & Increased Confidence	low-pressure, non-judgmental environment	9	16
	room for experimentation	7	
Immediate & Personalized Feedback	varied feedback	8	14
	suggestion of alternatives	6	
Engaging and Motivating Learning Experience	interactive exercises	6	14
	variety of activities	5	
	gamification elements	3	
Sum			66

Table 6 below presents a summary of the challenges posed by AI speaking partners as perceived by the participants.

Table 6.
Perceived Challenges of AI-Powered Chatbots for Oral Language Development

Theme	Subtheme	f	F (total)
Limitations in Natural Interaction	lack of human-like spontaneity	7	15
	difficulty with complex or nuanced conversations	4	
	limited adaptability to unpredictable input	4	
Technical and Usability Issues	internet connectivity dependence	7	12
	speech recognition errors	3	
	interface or navigation problems	2	
Over-Reliance on Technology	reduced chances of human authentic interaction	6	11
	difficulty transferring skills to real-world	5	
Concerns about Accuracy and Scope	occasional inaccurate feedback	4	7
	lack of socio-pragmatic sensitivity	3	
Hypercorrection	lack of human-like sensitivity in selective feedback	2	4
	limited user freedom in choosing feedback options	2	
Sum			49

As reflected in Tables 5 and 6 above, the Chatbot group generally demonstrated positive attitudes toward the approach, with a total of 66 mentions of advantages. However, they also voiced some concerns regarding the challenges in 49 instances. The learners particularly enjoyed the unprecedented access they experienced to a speaking partner, especially given the restrictions in an under-resourced EFL context, where oral practice opportunities beyond the classroom and exposure to authentic language are limited. They appreciated the fact that they could theoretically access the chatbot anytime, anywhere (9 mentions), and how this increased their chances of speaking practice (8 mentions), with no concerns about paying a tutor. In five cases, they also mentioned that the conversations seemed to adapt themselves to their preferences, knowledge, timing, and level. For instance, one participant emphasized that (the excerpts are translated into English):

> *As a student in Iran, finding native English speakers to practice with regularly is really difficult. With the Pi app, it's amazing because I can just open it on my phone whenever I have free time, like on the bus to university or before bed. It's like having an English tutor available 24/7, which is almost impossible to find.*

The students also reported a very positive effect on their anxiety levels and intimidation (16 instances). They felt free to experiment with language and test their hypotheses in a trial and error manner without the fear of being judged for their mistakes. They claimed that these features led to a gradual build-up of their confidence and trust, saving them from the embarrassment

which usually hinders their WTC in more traditional settings. One of the participants stated that:

> *I always felt nervous speaking in front of the teacher and other students. I was so worried about making mistakes and being judged. With Pi, it's completely different. I can relax and speak freely.*

In 14 cases, the students praised the availability of immediate, personalized, and varied feedback, which they often lack in their rather busy classes with time limitations in under-resourced educational settings. The participants also found the interactive and gamified nature of the app to be more stimulating than traditional learning methods, leading to increased enjoyment and engagement. This is particularly relevant in the context of this study and similar traditional contexts, where rote learning and textbook-based instruction, as the primary method of education, can sometimes lead to boredom and, eventually, demotivation. One learner reported that:

> *I like how Pi notices and reacts to changes in my English; it's a small thing, but it makes me feel good about my progress and makes me want to keep going. It is like a competition with myself.*

On the other hand, the learners also discussed the caveats they felt throughout their experience. Most importantly, they expressed concerns about the limitations of current technology in providing truly natural and fluid conversational experiences. Also, concerns were raised about how excessive interaction with a non-human agent can create unrealistic expectations in learners, distancing them from the facts of real-world communication. Specifically, some learners were skeptic about the capability of the app to expose them to a variety of culturally sensitive contexts and adhering to the socio-pragmatic norms governing a real interaction. Seven learners suggested that they had noticed the fabricated nature of the conversations, especially when it came to more complex issues and shifts in topic. For example, one learner stated that:

> *When you talk to a real person, they can react to anything you say, even if it's unexpected. Pi is good, but sometimes it feels like it's confused. If I say something a little off, it doesn't always follow.*

Another major group of concerns for participants, labelled as technical issues (F = 12), comprised frustrations with internet connection, app accuracy, and user experience. This especially makes sense in an under-resourced context,

where stable internet connection or global connectivity is not always accessible. The ASR also did not prove to be perfectly immaculate in their experience. This has probably occurred due to the training data and the limited exposure ASR services have had exposure to accents and pronunciations from particular contexts, such as Iran. Another group of learners with limited exposure to technology and restricted digital literacy, reported difficulties with the chatbot interface. This seems to be particularly true of under-resourced contexts, where general education does not yet extensively benefit from digitalized technologies.

Other than technical issues, a few learners also raised concerns about the possibility of forming a habit of overreliance on non-human training and interactions and failure to leave their comfort zones in the future for real-world communications. Although most respondents appreciated the low-pressure environment provided by chatbots (f = 9), some critically believed that this might falsely satisfy their need for authentic interactions for learning, hence, hindering the development of real-world conversational skills.

In four instances, learners reported occasional inaccurate feedback from the app. Also, three learners mentioned that despite the great help they were getting from the chatbot regarding their vocabulary and grammar, they still felt that attention to the social and cultural norms, such as politeness, was missing.

> *It doesn't tell me if I'm being too direct or too informal, which is important for me. You know, politeness is very important in our culture.*

Finally, a few learners felt frustrated with what they called overwhelming feedback from the application. They believed that their teachers were more sensitive in this regard and they had realized that instructors would sometimes choose to ignore some of their errors to keep the conversation flowing. Two learners also wished that the app features allowed them to select the degree of sensitivity, just like the difficulty settings in video games, so that they could decide for themselves the extent to which they wished to receive feedback and in which areas. Although, these kinds of adjustments were possible in *Pi* by providing the right prompt, it seems that some individuals' limited digital literacy stopped them from making the most out of the experience by properly personalizing the interactions.

Discussion

In order to discuss the results of the present study, it is necessary to first revisit the findings from both quantitative and qualitative phases in an integrated and holistic manner. The quantitative analyses generally confirmed that the integration of AI chatbots into L2 oral instruction in comparison with SCMC shows potential in contributing to oral proficiency development and significantly improves the self-efficacy of EFL learners in an under-resourced TESOL context. The qualitative findings, although focused on the AI chatbot group only, also showed that learners in this group generally felt comfortable with and enthusiastic about interactions with AI. This was mostly observed in their attitudes regarding chatbot's effect on their confidence, engagement, and motivation. Yet, concerns were also raised regarding the way AI could affect their learning behavior in the long run. Also, it must be noted that, a comparison with learners' attitudes in the SCMC group could have provided a more comprehensive picture.

The improvements observed in the quantitative phase of the study, can be attributed to several factors, perfectly matching with interactionist and sociocultural frameworks of second language learning. AI chatbot seems to have offered increased and immediate interaction and output opportunities, customized feedback, and adaptive learning environments, which cater to individual learner needs and promote active engagement while supporting their learning through tailored scaffolding (Guan et al., 2025; Kang & Sung, 2024). Personalized feedback in the AI group also allowed learners to identify, trace, and correct errors more systematically, leading to more effective learning and improved attitudes. Immediate interaction with AI tools provided learners with consistent practice opportunities, which are crucial for developing fluency and confidence in speaking. Successful autonomous technology-mediated language learning requires high levels of motivation and self-directed learning, which are dependent on one's confidence in their abilities to take learning into their own hands. The results confirm Teng's (2024) claim that in modern technology-based language learning, self-efficacy plays an important role in the relationship between social support and language learning anxiety.

The quantitative results of this study, corroborate a number of previous findings including those of Kang and Sung (2024), who found conversational chatbots positively influential in enhancing EFL learners' oral skills and self-directed learning as well as Guan et al. (2025) who found AI more effective than teacher-led conversational courses in boosting learners' speaking in an IDLE framework. Also, Huang (2024) confirmed that appropriately

employing ChatGPT, can lead to reduced speaking anxiety in language learners and encourage increased practice. As Liu and Zhang (2025) demonstrated, raising learners' literacy of new media and their awareness toward current educational technologies can boost their feelings of self-efficacy. Liu and Reinders (2025) also showed that AI chatbots can significantly improve learners' motivation which can in turn lead to boosted self-efficacy. The results also seem to supplement those of Huang and Mizumoto (2025), who had previously found a positive role for AI in improving learners' motivational self-system, leading to more positive self-efficacy beliefs. These findings also corroborate previous results in the Iranian context, which had found positive effects for AI chatbots including Speeko, Gliglish, Sayra, and Anima, on EFL learners' oral skills, willingness to communicate, and confidence (Baharloo & Miyan Baghi, 2024; Gholami et al., 2025; Shafiee Rad, 2024).

The qualitative findings confirm the practicality of using technology to enhance L2 learners' self-regulated learning as already suggested by Pan et al. (2025) and how AI, particularly chatbots, can offer personalized interactive support to help learners develop their autonomy and confidence in out-of-class learning situations. This supports Peng and Liang's (2025) findings, who showed the impact of AI interactions on EFL learners' WTC, technology acceptance, and attitudes. The results corroborate previous findings regarding the positive contributions of AI chatbots to language learning and elevating learners' psychological variables including WTC, foreign language enjoyment, attitudes, and reducing speaking anxiety (e.g., Liu et al., 2026; Wu et al., 2025; Yamaoka, 2024). The adaptability possibilities offered by chatbots was previously shown to contribute to L2 learners' self-regulated learning, which, in turn, supports lifelong, autonomous learning.

However, despite the generally positive attitudes, drawbacks were also mentioned. Learners' expressed concerns about the authenticity and naturalness of the AI-led conversation, as well as their critical outlooks on creativity building, especially in real-world situations, echo the arguments of Tin (2025), who has recently warned that:

> *Improper use of AI tools such as ChatGPT can result in collective novelty decay and deprive us of valuable cognitive opportunities to reflect on, draw from, and reshape our personal, emotional and socio-cultural experiences and to engage in unprompted creativity when performing a language task. (p. 46)*

Therefore, the findings emphasize that although some previous research, specifically, Shin & Lee (2024) found agreeable attitudes toward the authenticity, context-appropriateness, and realistic qualities of *Pi* in comparison with other AI-based platforms, these features still have considerable room for improvement to fully meet the needs of their users. Learners' need for more control on their learning experience with AI-powered chatbots is in line with Chong and Reinders' (2025) conceptualization of learner autonomy and the greater role which needs to be assigned to them for better results. As Wiboolyasarin et al.'s (2025) systematic review revealed, despite the overall efficacy of AI chatbots in supporting most language skills, a lack of authentic capability in creating deep contextual interactions was reported and it seems that they lagged behind human interactions in emotional engagement with the learners. According to Koç and Savaş (2025), "the unnaturalness of AI chatbot-human interactions" was already a common challenge reported in at least nine prior studies. Godwin-Jones (2024) argued, "due to how AI systems understand human language, based on a mathematical model using statistical probability, they lack the lived experience to be able to use language with the same social awareness as humans" (p. 1). The findings seem to corroborate Godwin-Jones (2021) in the necessity of taking "sociocultural/pragmatic" aspects of technology-enhanced language learning into consideration and giving room to the "organic interplay between learner and environment" (p. 6).

Therefore, learners' comments on cultural aspects in AI-mediated interactions in the present study, seem to reflect well-justified and timely concerns. Hence, it seems that despite the great potential it is carrying, AI chatbots still have a long way to go in order to overcome these shortcomings, especially given the fact that the geographical distribution of training data for large language models does not currently represent the world population and largely misses out on under-resourced and hence under-represented TESOL contexts.

Conclusion and Implications

From an interactionist perspective, the adoption of AI tools in order to enhance L2 oral proficiency and self-efficacy in under-resourced contexts offers several benefits and challenges, which require more in-depth, context-based investigations. The present study showed that AI offers promising perspectives in overcoming these barriers to enhance learning outcomes of EFL students and push language education to more equitable and sustainable standards around the globe. However, as Fryer et al. (2020) had also observed,

the efficacy of chatbots, including their implications for autonomous and self-directed learning, can be improved if learners' attitudes, experiences, and needs are taken into consideration.

AI tools can be a valuable supplement to more traditional teaching methods, especially in contexts with limited access and resources, providing learners with additional support and interactive experiences, enhancing their language skills and sense of self-efficacy. While AI can provide practice opportunities, it might still fall short of replicating the nuanced, context-rich interactions that human educators offer. This limitation points to the importance of a balanced approach in which AI serves as an adjunct, not a replacement, to other teaching methods. It is crucial that future research addresses the challenges and limitations associated with AI in language learning, such as technical difficulties and the need for human interaction, to maximize its effectiveness. The human element in language education, where cultural intricacies, emotional intelligence, and spontaneous interaction play pivotal roles, seems to remain irreplaceable; yet, AI tools offer several advantages, including enhanced practice opportunities, reduced anxiety, immediate feedback, and an engaging learning experience, which can contribute to improved speaking skills and increased confidence in learners' English language abilities.

These findings can offer several implications for language learners, educators, and curriculum developers, especially in under-resourced TESOL contexts. Most importantly, teachers can encourage the ethical and responsible use of AI chatbots, not just as sources of information, but also as speaking partners and tutors. EFL teachers, if informed and trained properly, can model the use of AI chatbots as speaking partners for their learners and get their learners started on this technology by offering gradually declining monitoring. The support can continue as long as the learners form realistic expectations of AI and learn how to adjust their learning behaviors and feelings in order to maximize the benefits they reap from such interactions. The results also point out the limitations and shortcomings of such tools, which might help learners and teachers in preparing to compensate for these caveats and making optimal use of their potentials instead of getting frustrated by their hurdles.

Curriculum developers can also integrate awareness-raising, instructions and trainings, as well as supplementary tasks and activities to promote AI-supported oral practice and autonomous learning beyond the classroom. Being informed of the possible learning benefits and attitudinal effects on learners and their sense of self-efficacy can help both curriculum developers

and teacher educators design and implement more accurate and accommodating tasks and activities in their practices.

A number of limitations are associated with the present study. First of all, lack of access to other under-resourced contexts delimited the scope of this study to the case of Iran. Future research can collect and compare data from several such settings and offer a broader perspective. Also, the number of participants was rather limited and digital literacy and technology acceptance were not factored in. Regarding sample limitations, only two experimental groups were recruited in the present study and no control groups, featuring traditional face-to-face interaction, were integrated. Of course, it can be argued that given the well-established role of interaction in developing L2 speaking skills, the control group was dismissed; however, future research can replicate the design with an additional control group, especially to track learners' experience during the treatment. Furthermore, the qualitative data collection was restricted to self-reports through semi-structured interviews, while a triangulation of this data with introspective or retrospective verbal protocols as well as learner diaries can enrich the findings of future studies. Also, due to space limitations, the attitudes and experiences of the learners in the SCMC group were not observed in the present study. The comparison between AI and SCMC group data regarding their attitudes could have enriched the qualitative analysis of the study by providing a more comprehensive picture. Finally, due to the restrictions experienced in an under-resourced context, a single conversational bot was employed, while, the rapid progress in AI chatbots encourages the use of up-to-date and pioneering apps and platforms for further studies.

To continue this line of research, future endeavors can also venture into comparative studies to understand how cultural factors in different contexts influence the effectiveness and acceptance of AI tools in language learning, paving the way for culturally adaptive AI solutions. Subskills of oral proficiency can also be explored in more detail in future research. There is also a need for research focusing on the ethical implications, privacy issues, and data security when using AI in education. This would include studying how learners and educators perceive and manage the risks associated with AI tools, leading to guidelines for ethical AI use in educational settings.

Disclosure Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be considered as a potential conflict of interest.

Funding

No funding was received for this research project.

Acknowledgements

The authors would like to appreciate the cooperation of all participants who agreed to take part in this study and the IELTS instructors who cooperated in the rating process.

ORCID

Mohammad Mahdi Hajmalek https://orcid.org/0000-0003-3688-6029
Negin Etedal https://orcid.org/0009-0007-3160-3452

References

Abrams, Z. I. (2003). The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal*, *87*(2), 157–167. https://doi.org/10.1111/1540-4781.00184

Baharloo, A., & Miyan Baghi, A. (2024). The impact of AI-assisted learning on EFL speaking skills: A mixed-methods study in the Iranian context. *Technology Assisted Language Education*, *2*(4), 69–96. https://doi.org/10.22126/tale.2025.11299.1070

Bailey, K. M., & Christian, D. (Eds.). (2021). *Research on teaching and learning English in under-resourced contexts*. Routledge. https://doi.org/10.4324/9781003057284

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191–15. https://psycnet.apa.org/doi/10.1037/0033-295X.84.2.191

Chong, S. W., & Reinders, H. (2025). Autonomy of English language learners: A scoping review of research and practice. *Language Teaching Research*, *29*(2), 607–632. https://doi.org/10.1177/13621688221075812

Curtis, A. (2021). What do we mean by under-resourced context? In K. M. Bailey & D. Christian (Eds.), *Research on teaching and learning English in under-resourced contexts* (pp. 14–28). Routledge.

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*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Du, Y., & Reynolds, B. L. (2025). Rooted in and beyond interaction: A systematic review of interactive affordances of chatbots for language learning amidst the rise of large language models. *Acta Psychologica*, *259*, 105307. https://doi.org/10.1016/j.actpsy.2025.105307

Fryer, L. K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, *24*(2), 8–22. Retrieved from http://hdl.handle.net/10125/44719

Ghiasvand, F., Kogani, M., & Alipoor, A. (2024). "I'm not ready for this metamorphosis": An ecological approach to Iranian and Italian EFL teachers' readiness for artificial intelligence-mediated instruction. *Teaching English with Technology*, *24*(3), 18–40. https://doi.org/10.56297/vaca6841//BFFO7057/ISKI2001

Gholami, A., Ahmadi, S., & Taheri, H. (2025). The effect of using chatbots and digital literacy on Iranian EFL learners' speaking: A mixed methods study. *Journal of Language Horizons*, *9*(3), 7–31. https://doi.org/10.22051/lghor.2024.46124.1893

Godwin-Jones, R. (2021). Evolving technologies for language learning. *Language Learning & Technology*, *25*(3), 6–26. http://hdl.handle.net/10125/73443

Godwin-Jones, R. (2024). Generative AI, pragmatics, and authenticity in second language learning. *arXiv*, 2410–14395. https://doi.org/10.48550/arXiv.2410.14395

Godwin-Jones, R., O'Neill, E., & Ranalli, J. (2024). Integrating AI tools into instructed second language acquisition. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 9–23). Iowa State University Digital Press. https://doi.org/10.31274/isudp.2024.154.02

Guan, L., Zhang, E. Y., & Gu, M. M. (2025). Examining generative AI–mediated informal digital learning of English practices with social cognitive theory: A mixed-methods study. *ReCALL*, *37*(3), 315–331. https://doi.org/10.1017/S0958344024000259

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, *25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Healey, D., Hanson-Smith, E., Hubbard, P., Ioannou-Georgiou, S., Kessler, G., & Ware, P. (2011). *TESOL technology standards: Description, implementation, integration*. TESOL. https://doi.org/10.1002/9781118784235.eelt0453

Hedayati, H. F., & Marandi, S. S. (2014). Iranian EFL teachers' perceptions of the difficulties of implementing CALL. *ReCALL*, *26*(3), 298–314. https://doi.org/10.1017/S0958344014000172

Huang, J. (2024). Enhancing EFL speaking feedback with ChatGPT's voice prompts. *International Journal of TESOL Studies*, *6*(3), 4–23. https://doi.org/10.58304/ijts.20240302

Huang, J., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the L2 motivational self-system. *Education and Information Technologies*, *30*(5), 6435–6454. https://doi.org/10.1007/s10639-024-13071-6

Hubbard, P. (2021). Revisiting the TESOL technology standards for teachers. *CALICO journal*, *38*(3), 319-337. https://doi.org/10.1558/cj.20068

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1-2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, *55*(1), 48–63. https://doi.org/10.1080/15391523.2022.2142873

Jiang, Y., & Chun, D. (2023). Web-based intonation training helps improve ESL and EFL Chinese students' oral speech. *Computer Assisted Language Learning*, *36*(3), 457–485. https://doi.org/10.1080/09588221.2021.1931342

Joo, S. H. (2024). Generative AI as writing or speaking partners in L2 learning: Implications for learning-oriented assessments. Studies in Applied Linguistics & TESOL at Teachers College, Columbia University, *24*(1), 54–9 https://doi.org/10.7916/h6v0-my25

Kang, S., & Sung, M. (2024). EFL students' self-directed learning of conversation skills with AI chatbots. *Language Learning & Technology*, *28*(1), 1–19. https://hdl.handle.net/10125/73600

Karatay, Y., & Hegelheimer, V. (2021). CALL teacher training—considerations for low-resource environments. *CALICO journal*, *38*(3), 271–295. https://doi.org/10.1558/cj.20159

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. *Language Learning & Technology*, *26*(1), 1–18. http://hdl.handle.net/10125/73463

Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, *37*(1), 4–21. https://doi.org/10.1017/S0958344024000168

Lai, C., & Sundqvist, P. (2025). Research agenda: Synergizing in-class and out-of-class language learning with technology. *Language Teaching*, 1–22. https://doi.org/10.1017/S026144482500014X

Lantolf, J. P. (Ed.). (2000). *Sociocultural theory and second language learning* (Vol. 78, No. 4). Oxford University Press. https://doi.org/10.1017/S0267190500002646

Leeming, P., Vitta, J. P., Hiver, P., Hicks, D., McLean, S., & Nicklin, C. (2024). Willingness to communicate, speaking self-efficacy, and perceived communicative competence as predictors of second language spoken task production. *Language Learning*, *74*(4), 917–949. https://doi.org/10.1111/lang.12640

Lin, C. C., Huang, A. Y., & Lu, O. H. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review. *Smart Learning Environments*, *10*(1), 41. https://doi.org/10.1186/s40561-023-00260-y

Lin, H. (2015). Computer-mediated communication (CMC) in L2 oral proficiency development: A meta-analysis. *ReCALL*, *27*(3), 261–287. https://doi.org/10.1017/S095834401400041X

Liu, M., & Reinders, H. (2025). Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. *System*, *128*, 103544. https://doi.org/10.1016/j.system.2024.103544

Liu, X., & Zhang, D. (2025). Learning critically and confidently: A correlation study of the new media literacy and English learning self-efficacy of junior high school students. *International Journal of TESOL Studies*, *7*(1), 88–104. https://doi.org/10.58304/ijts.20250106

Liu, Z., Zhang, W., & Yang, P. (2026). Can AI chatbots effectively improve EFL learners' learning effects? A meta-analysis of empirical research from 2022–2024. *Computer Assisted Language Learning*, *39*(4), 1176–1202. https://doi.org/10.1080/09588221.2025.2456512

Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input and second language acquisition* (pp. 377–393). Newbury House.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413– 468). Academic Press.

Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies. *International Journal of Applied Linguistics*, *35*(2), 834–851. https://doi.org/10.1111/ijal.12668

Moslemi Nezhad Arani, S., Zarei, A. A., & Sarani, A. (2024). Factors impeding implementing CALL-based instruction: A mixed-methods study. *System*, *126*, 103461. https://doi.org/10.1016/j.system.2024.103461

Pan, M., Lai, C., & Guo, K. (2025). Technology-enhanced self-regulated second language learning interventions: A systematic review. *System*, *131*, 103675. https://doi.org/10.1016/j.system.2025.103675

Peng, J. E., & Liang, W. (2025). Willingness to communicate with artificial intelligence (AI)? Insights from tracking EFL learners' perceived acceptance and chat output. *Computer Assisted Language Learning*, 1–27. https://doi.org/10.1080/09588221.2025.2486147

Shabbir, A., Rizvi, S., Alam, M. M., & Su'ud, M. M. (2024). Beyond boundaries: Navigating the positive potential of ChatGPT, empowering education in underdeveloped corners of the world. *Heliyon*, *10*(16), 1–13. https://doi.org/10.1016/j.heliyon.2024.e35845

Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of Speeko application. *Innovation in Language Learning and Teaching*, *18*(4), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Shin, D., & Lee, J. H. (2024). Moving from off-the-shelf chatbots to a user-designed bespoke L2 chatbot. *Language Learning & Technology*, *28*(1), 1–14. https://hdl.handle.net/10125/73581

Smith, B. (2003). Computer–mediated negotiated interaction: An expanded model. *The Modern Language Journal*, *87*(1), 38–57. https://doi.org/10.1111/1540-4781.00177

Suryanarayana, K. S., Kandi, V. P., Pavani, G., Rao, A. S., Rout, S., & Krishna, T. S. R. (2024). Artificial intelligence enhanced digital learning for the sustainability of education management system. *The Journal of High Technology Management Research*, *35*(2), 100495. https://doi.org/10.1016/j.hitech.2024.100495

Teng, M. F. (2024). Do self-efficacy belief and emotional adjustment matter for social support and anxiety in online English learning in the digital era? *Digital Applied Linguistics*, *1*, 2227–2247. https://doi.org/10.29140/dal.v1.2227

Teng, M. F. (2025). Metacognition, self-efficacy belief, language learning motivation, and perceived progress in online English learning: A cross-lagged analysis. *International Journal of TESOL Studies*, *7*(1), 4–29. https://doi.org/10.58304/ijts.20250102

Teng, M. F., Wang, C., & Wu, J. G. (2023). Metacognitive strategies, language learning motivation, self-efficacy belief, and English achievement during remote learning: A structural equation modelling approach. *RELC Journal*, *54*(3), 648–666. https://doi.org/10.1177/00336882211040268

Tin, T. B. (2025). Creativity in the AI era: Humans vs. ChatGPT in unprompted playful language tasks. *International Journal of TESOL Studies*, *7*(2), 46–65. https://doi.org/10.58304/ijts.250125

Todd, R. W. (2025). Generative AI as a disrupter of language education. *International Journal of TESOL Studies*, *7*(2), 37–45. https://doi.org/10.58304/ijts.250127

Wang, C., Kim, D. H., Bong, M., & Ahn, H. S. (2013). Examining measurement properties of an English self-efficacy scale for English language learners in Korea. *International Journal of Educational Research*, *59*, 24–34. https://doi.org/10.1016/j.ijer.2013.02.004

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand*, *14*, 100224. https://doi.org/10.1016/j.amper.2025.100224

Wu, T. T., Hapsari, I. P., & Huang, Y. M. (2025). Effects of incorporating AI chatbots into think–pair–share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 1–39. https://doi.org/10.1080/09588221.2025.2478271

Yamaoka, K. (2024). ChatGPT's motivational effects on Japanese university EFL learners: A qualitative analysis. *International Journal of TESOL Studies*, *6*(3), 24–35. https://doi.org/10.58304/ijts.20240303

Yang, L., & Li, R. (2024). ChatGPT for L2 learning: Current status and implications. *System*, *124*, 103351. https://doi.org/10.1016/j.system.2024.103351

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, *38*(3), 553–586. https://doi.org/10.1017/S027226311500025X

Appendix A. Semi-Structured Interview

1. During the past semester, you have been practicing your speaking with *Pi*; did you like this experience?
2. How would you describe your overall experience using the *Pi* application?
3. What did you like about it most?
4. What didn't you like about using *Pi*?
5. Do you think using *Pi* has improved your skills in speaking English? In what ways?
6. How do you feel about your ability to speak English now compared to when you started using *Pi*?
7. Do you think using *Pi* has affected your confidence to use or speak English? In what ways?
8. Did you experience any challenges or difficulties when speaking with *Pi*? How did you overcome them?
9. Would you recommend using AI applications like *Pi* to other English learners? Why or why not?
10. What do you think your teacher could do to improve your experience of using artificial intelligence for practicing English?