Journal of English Language Teaching and Learning

University of Tabriz

Volume 17, Issue 35, 2025

Exploring the Effects of Chatbot-Assisted Language Learning on Learners' Language Achievement and Sense of Metavolvement

Mohammad-Javad Gholami-Hossein-Abad

Ph.D. Candidate, Department of English, Ferdowsi University of Mashhad, Iran. E-mail: mjgholamihosseinabad@alumni.um.ac.ir

Reza Pishghadam (Corresponding Author)

Professor, Department of English, Ferdowsi University of Mashhad, Iran. E-mail: pishghadam@um.ac.ir

Zargham Ghabanchi

Associate Professor, Department of English, Ferdowsi University of Mashhad, Iran. E-mail: ghabanchi@um.ac.ir

Elham Naji-Meidani

Assistant Professor, Department of English, Ferdowsi University of Mashhad, Iran. E-mail: elhanaji@um.ac.ir

ARTICLE INFO:

Received date:
2025.05.30
Accepted date:
2025.06.30

Print ISSN: 2251-7995
Online ISSN: 2676-6876

Keywords:

AI, Artificial Intelligence, Emotioncy, English Language Learning, Language Achievement

Abstract

Artificial intelligence (AI) is changing language education through personalized and interactive practice. However, its impact on language achievement and emotional aspects of learning is unclear. Therefore, this mixed-methods study investigated the effects of Chatbot-Assisted Language Learning (CHALL) on learners' English language achievement and sense of metavolvement—the deepest level of metacognitive engagement. In phase 1, the Sense of Metavolvement Scale (SOMS) was developed and validated. In phase 2, following a quasi-experimental design, 44 intermediate Iranian EFL learners were divided into an experimental group (n=22, using the AI chatbot Pi) and a control group (n=22, receiving conventional classroom instruction). Both groups had pre- and post-tests for language achievement and metavolvement, and the experimental group sat for semi-structured interviews post-intervention. There were significant gains in the experimental group's in-class metavolvement and language achievement. There were also significant differences between the two groups in post-test metavolvement, while there was no significant difference in their post-test language achievement. Furthermore, interview data illustrated CHALL as a low-anxiety practice context with cultural and pragmatic limitations. Therefore, CHALL can complement but not replace conventional instruction.

Citation: Gholami-Hossein-Abad, M.-J.; Pishghadam, R.; Ghabanchi, Z. & Naji-Meidani, E. (2025). Exploring the Effects of Chatbot-Assisted Language Learning on Learners' Language Achievement and Sense of Metavolvement. *Journal of English Language Teaching and Learning*, 17 (35), 245-266. DOI: 10.22034/elt.2025.67757.2795

Introduction

Artificial Intelligence (AI) is revolutionizing language learning and teaching through interactive and personalized tools, which can potentially improve learning outcomes and, among these tools, AI chatbots have gained particular attention (Gholami & Al Abdwani, 2024). They allow learners to practice language in natural, dialogic interactions through simulated human-like conversations and provide them with personalized feedback (Zhai & Wibowo, 2023). Given the current global demand for accessible, individualized instruction outside traditional classrooms, Chatbot-Assisted Language Learning (CHALL) holds great promise. However, previous studies have mainly investigated its potential in limited areas of language competencies, such as vocabulary or grammar, overlooking the impacts of CHALL on learners' overall language proficiency and metacognitive engagement with the learning process, what Pishghadam et al. (2019a) termed metavolvement). It is also unclear how cultural and pedagogical contexts shape learners' perceptions and use of AI chatbots (Wu & Li, 2024), which is especially important in learning environments such as the Iranian EFL context, where educational traditions and cultural norms may impact learners' engagement with interactive technology. In such contexts, pedagogical affordances of CHALL may not be fully realized if they do not align with learners' cultural expectations and classroom practices (Wu & Li, 2024). Therefore, to leverage the affordances of CHALL in language education, it is essential to understand how it impacts learners' language achievement and their sense of metavolvement (as a representation of their metacognitive engagement) in these environments and how learners perceive and interact with AI chatbots. To address these gaps, this study aims to investigate the effects of CHALL on Iranian EFL learners' language achievement and their sense of metavolvement using a mixed-methods approach. This study is guided by the following research questions:

1. Does the Sense of Metavolvement Scale (SOMS) demonstrate robust psychometric properties for assessing learner engagement?
2. Does CHALL significantly enhance learners' overall English language achievement?
3. Does CHALL significantly improve learners' sense of metavolvement?
4. How do learners perceive CHALL compared to conventional classroom instruction?

1. Literature review

1.1. Evolution of Technology-Assisted Language Acquisition: From CALL to CHALL

Over the past decades, the use of technology in language learning contexts has vastly developed. Computer-assisted language learning (CALL) was the earliest form of technology-assisted language learning developed in the 1960s, which was limited to grammar drills and vocabulary exercises provided by computers (Warschauer & Healey, 1998). In the 1980s and 1990s, CALL was transformed through incorporating multimedia, hypertext, previously inactive media such as CD-ROMs, and artificial intelligence (AI) at a very basic level (Chapelle, 2001). Subsequent to CALL, Mobile-Assisted Language Learning (MALL) paved the way for portable language learning and practice through integrated activities which could more deeply engage learners (Stockwell, 2010; Viberg & Grönlund, 2013). More recently,

Robot-Assisted Language Learning (RALL) introduced language robots as embodied conversational agents to bring about socio-emotional engagement (Belpaeme et al., 2018). The latest form of technology-assisted language learning, CHALL, exploits natural language processing (NLP) and machine-learning algorithms to create personalized, interactive, and adaptive learning experiences (Huang et al., 2023; Zhai & Wibowo, 2023), allowing a more natural development of language skills through interaction (Zhang et al., 2023).

1.2. Theoretical Underpinnings and Limitations of CHALL

CHALL is consistent with the core theories of second language acquisition (SLA). Consistent with the Interaction Hypothesis (Long, 1996), which emphasizes interaction as a tool to provide comprehensible input and opportunities for meaning negotiation, CHALL seeks to simulate conversation and help learners negotiate meaning through interaction (Golonka et al., 2014). Furthermore, in line with the Corrective Feedback Theory (Lyster & Saito, 2010), CHALL provides corrective feedback on writing, pronunciation, and lexical choice instantly (Yang et al., 2022). In line with Vygotsky's (1978) Sociocultural Theory, CHALL can be considered a "more knowledgeable other", which interacts with learners, places them in their Zone of Proximal Development (ZPD) through its scaffolding ability, and facilitates their communicative growth (Bernstein et al., 2010).

Despite these strengths, CHALL has some limitations as well. Firstly, AI chatbots do not have the affordances of a physical interlocutor, such as the ability to interact flexibly and negotiate meaning (Huang et al., 2022). Furthermore, chatbots only provide feedback on surface-level errors and may overlook deeper pragmatic and discourse-related ones (Kim et al., 2021). Learners have also shown different perceptions toward CHALL. While some enjoy the neutral, non-judgmental feedback provided by AI chatbots (Fryer et al., 2020), others dislike their engagement with CHALL, referring to it as repetitive, pre-scripted, and emotionally disconnected (Fryer et al., 2020).

1.3. The Impact of CHALL on Language Achievement

Recent studies have demonstrated that CHALL can positively impact various language skill areas. Regarding speaking ability, Kim et al. (2021) found that chatbot practice improves fluency, pronunciation accuracy, and confidence by providing a low-anxiety atmosphere. However, chatbots have difficulty recognizing and correcting more complex errors related to prosody or intonation (Yang et al., 2021). As for listening, research has shown that exposure to chatbots' naturally varied accents and speech rates leads to improvements in listening skills and listening comprehension (Yang et al., 2022). On the other hand, some studies suggest that the limited variations in voice and real-life conversation features lead to chatbots' poor representation of actual communication (Duong & Suppasetseree, 2024).

CHALL also proves to be beneficial for reading comprehension. Students using chatbots for adaptive reading tasks demonstrated improvements in reading comprehension and self-efficacy (Nagata et al., 2020). However, chatbots cannot promote deeper-level inferential skills needed for the comprehension of more complex texts (Panda & Kaur, 2023). Regarding writing skills, CHALL was found to improve grammatical accuracy, coherence, syntactic variety, and lexical sophistication (Zhang et al., 2023). Nevertheless, the feedback was often confined to sentence-

level corrections, neglecting higher-order writing aspects, such as argument development or coherence across paragraphs (Panda & Kaur, 2023).

In terms of grammar and vocabulary acquisition, CHALL has also shown potential positive impacts. Qasem et al. (2023) found that learners using chatbots for grammar practice retained grammatical rules better. AI chatbots were also found to aid in vocabulary expansion and retention by furnishing contextual word definitions, synonyms, and examples (Qasem et al., 2023). However, the absence of discourse-level error correction and occasional misinterpretations by chatbots present limitations (Wu & Li, 2024). What is evident is that using CHALL together with conventional methods would ensure maximum benefit from AI (Zhang et al., 2023).

1.4. The Emotioncy Model and CHALL Integration

The emotioncy model, developed by Pishghadam (2016), links sensory and emotional engagement to language learning and posits that greater emotional and sensory connections to the learning material foster better retention and processing (Pishghadam et al., 2017; Pishghadam et al., 2019b).

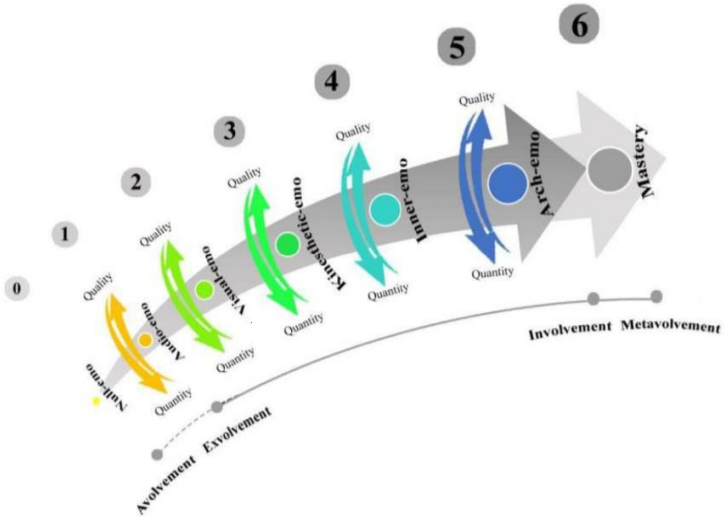

Figure 1. *Emotioncy Levels, adapted from Pishghadam et al. (2019a)*

As can be seen in Figure 1, emotioncy levels proceed in a hierarchical order from avolvement (no previous exposure to the concept) to exvolvement (auditory, visual, or kinesthetic exposure to the concept) and involvement, which includes inner emotioncy (deep personal experience with the concept) and arch emotioncy (having done research on the concept). Pishghadam et al. (2019a) added another level to the Emotioncy Model called metavolvement, which is an intense level of engagement with the concept wherein learners are so involved that they invent their own knowledge and take control of their learning. The higher the level of sensory-emotional involvement (i.e., emotioncy) with the learning process, the better the learning outcomes (Pishghadam & Shayesteh, 2017).

Since CHALL is multimodal, it has the potential to promote lower levels of emotioncy, like exvolvement, by exposing learners to sensory involvement through visual, auditory, or even kinesthetic components (Pishghadam et al., 2023a). However, chatbot systems may come short

in providing real-life contextual experiences that build toward arch or inner emotioncy, such as creating immersive learning environments with physical interaction (Pishghadam et al., 2023b; Zhang et al., 2023).

1.5. Learners' Perceptions of CHALL

Research has shown that learners appreciate CHALL's anxiety-free and instant feedback, which, in line with Krashen's (1989) affective filter hypothesis, fosters language learning (Kim et al., 2021; Huang et al., 2023). Learners who are normally more reserved in a face-to-face context, such as Iranian EFL learners, are possibly more inclined to favor CHALL, and the artificially low-stakes, AI-powered experience can promote their level of engagement (Wei, 2023). On the other hand, in some studies, learners have expressed that sometimes chatbots disappoint them by quickly ending conversations or not following culturally specific linguistic norms, thereby failing to match their cultural expectations (Fryer et al., 2020). Misunderstandings with idiomatic expressions have also affected students' perceptions of interaction with a chatbot (Wei, 2023).

2. Methodology

2.1. Research Design

This mixed-methods study included two distinct phases. The first phase involved a development and validation process of the Sense of Metavolvement Scale (SOMS) and the second phase utilized a quasi-experimental design with 44 intermediate level learners (experimental, use of a chatbot Pi; control, conventional instruction). Quantitative data consisted of pre/post-tests and SOMS measures with qualitative data obtained from interviews.

2.2. Participants

The participants of phase one of the study, which involved the development and validation of the Sense of Metavolvement Scale (SOMS), were selected through convenience sampling from several private language academies located in Kerman and Rafsanjan, Iran.

Table 1. *Participant Demographics – Phase 1*

Characteristic	Description
Sample Size	213 EFL learners
Gender	Female: 177 (83.1%) Male: 36 (16.9%)
Age Range	11-53 (M = 18.2, SD = 6.28)
English Proficiency	CEFR B1 to C1
Language Learning Duration	<1 year: 6 1 year: 33 2 years: 32 3 years: 13 >3 years: 129
Degree	Pre-diploma: 156 Diploma: 9 Associate: 4 Bachelor: 30 Master: 14
First Language (L1) and Nationality	Persian (all Iranian)

The final sample included 213 EFL learners (177 females [83.1%] and 36 males [16.9%]) whose ages ranged from 11 to 53 years (M = 18.2, SD = 6.28), and their educational qualifications varied from pre-diploma to diploma, bachelor's degree, and master's degree. Their proficiency level ranged from B1 to C1, and they all had studied English for a varying duration of 1 to 3 years. Although there was an imbalance in gender, it was consistent with our conceptualization of metavolvement as a cognitive construct which, unlike personality traits, is not influenced by gender characteristics.

Participants of phase two were 44 (29 female and 15 male) Iranian EFL learners studying English in intermediate courses at Mahan Asa Language Academy in Kerman, Iran (a private institution with a high fee structure, primarily attracting participants from a middle-to-high socioeconomic background).

Table 2. *Participant Demographics – Phase 2*

Characteristic	Total Sample (N = 44)	Experimental (N = 22)	Control (N = 22)
Gender	Female: 29 (65.9%) Male: 15 (34.1%)	Female: 13 (59%) Male: 9 (41%)	Female: 16 (72.7%) Male: 6 (27.3%)
Age Range	16-46 (M = 29.75, SD = 8.09)	17-45 (M = 29.18, SD = 7.45)	16-46 (M = 30.32, SD = 8.82)
English Proficiency	CEFR B1 (intermediate)	CEFR B1 (intermediate)	CEFR B1 (intermediate)
Language Learning Duration	<1 year: 8 1 year: 7 2 years: 12 3 years: 3 >3 years: 14	Similar distribution	Similar distribution
Degree	Pre-diploma: 1 Diploma: 7 Bachelor: 16 Master: 17 PhD: 3	Similar distribution	Similar distribution
First Language (L1)	Persian (all Iranian)	Persian (all Iranian)	Persian (all Iranian)

This group of participants were also selected through convenience sampling based on their enrollment in the academy's intermediate-level program, denoting CEFR B1. We selected intermediate students as we were sure they were able to engage meaningfully with the chatbot and benefit from the potentially reciprocal aspect of input and corrective feedback. The participants' age range was 16 to 46 (M = 29.75, SD = 8.09), encompassing teenager and adult learners, and their educational qualifications ranged from pre-diploma to PhD. We also ensured that the participants had no experience using a quality chatbot for language learning so that their prior familiarity would not impact the findings. The socioeconomic homogeneity of the sample, which was partially confirmed based on the fee structure of the private institute, provided some control over confounding variables regarding educational experiences and access to technology.

2.2. Instruments

2.2.1. The Textbook: American English File 3 (Third Edition)

American English File 3 (Latham-Koenig et al., 2020) was the coursebook used in phase 2 as the teaching material for both groups. This coursebook contains a strategic balance of reading, writing, listening, and speaking with grammar, vocabulary, and pronunciation practice. The material is grounded in real-world situations, has media built into the chapters, and contains entertaining interactive activities.

2.2.2. Language Proficiency Tests

Pre- and post-tests were used to determine learners' language achievement in terms of language skill (reading, writing, speaking, and listening) and key language components (grammar, vocabulary, and pronunciation). The pre- and post-tests were both in-house, institution-wide proficiency tests based selectively on the American English File 3 series.

2.2.3. Sense of Metavolvement Scale (SOMS)

The Sense of Metavolvement Scale, which was developed and validated in this study, was administered before and after the intervention to assess changes in students' sense of metavolvement following the use of CHALL. The scale consists of two sub-constructs of in-class and out-of-class metavolvement (6 items each). The present study also confirmed the reliability of the SOMS ($\alpha = .90$).

2.2.4. Chatbot Intervention: Pi

The intervention was conducted in the present study by incorporating the Pi chatbot into the teaching-learning process for the experimental group as a supplement to conventional classroom instruction. Pi was selected for this study due to its advanced conversational engagement skills and its capacity to provide personalized adaptive feedback. Pi gives real-time multimodal input (text or voice) to students, provides answers suitable to particular contexts, personalizes the lessons, and provides contextual information, personalized scaffolding based on individual learners' language responses, and real-time feedback. One of the characteristic features of Pi is the low-anxiety atmosphere it creates through tone adjustments in giving feedback and its encouragement-based approach to reinforce instruction.

2.3. Procedure

Phase 1 of the study focused on developing and validating the SOMS. The first item panel of the scale was made based on the Emotioncy Model. Two experts in applied linguistics and psychometrics were recruited to help address item clarity and content validity, leading to the removal of four ambiguous items. Following this, a pilot study involving 13 participants, a random sample of the 213 EFL learners selected for phase 1, was conducted to confirm readability, which also led to making some minor changes to the scale. The developed scale was then administered to the 213 participants. After that, the reliability of the scale was checked through an analysis of Cronbach's alpha coefficients, and Confirmatory Factor Analysis (CFA) was run to establish the construct validity of the scale.

Phase 2 was a combination of quantitative (quasi-experimental design) and qualitative data collection. In accordance with the quasi-experimental design, quantitative data were collected

through administering a test of language achievement and the SOMS before and after the intervention (pre- and post-tests) to examine the effects of CHALL on students' language achievement and sense of metavolvement, respectively. Both groups' final grades in the previous term were considered as their pre-test scores of language achievement, establishing baseline language competence with a high degree of reliability. Both groups then participated in the new term, which lasted five weeks, and attended six 90-minute weekly sessions at Mahan Asa Language Academy, Kerman, Iran. The first researcher provided instruction to both groups. The control group was instructed through the *American English File 3* coursebook (Latham-Koenig et al., 2020) using conventional classroom instruction methods, such as demonstrating and explaining vocabulary and grammar, completing coursebook exercises, and doing role-plays. The experimental group also received the same conventional classroom instruction, supplemented by the use of the Pi chatbot in and out of class.

In-class CHALL activities for the experimental group started with a 5- to 10-minute warm-up dialogue, where they were given the topic for that session's lesson and were asked to talk about it with Pi using their cell phones. This dialogue helped to activate their background knowledge about the topic, or in cases of unfamiliar topics, create a primary level of familiarity with the topic. During the session, Pi served as an assistant for each student. More specifically, after the instructor taught grammar or vocabulary, students asked Pi to briefly review the lesson for them and provide more examples and exercises while adapting to their individual proficiency level. This way, each student had some extra time to grasp and practice the lesson at their own pace and level. Students also participated in role-plays within lessons that reflected the coursebook scenarios. For example, students practiced role-plays for social obligations in formal settings (Unit 4) with Pi, and based on their performance, Pi provided extra practice for them.

Students also used Pi for out-of-class activities. Depending on the covered lesson and the task at hand, students requested Pi to provide grammar, vocabulary, or pronunciation exercises related to the lesson. If they had a writing assignment, they used Pi for brainstorming their ideas and receiving feedback on grammar, vocabulary, and the whole writing. They also asked Pi to generate reading texts similar to the ones in their coursebook while adapting it to their level for further practice. They also engaged in 30 minutes of daily conversational practice with Pi, where Pi checked their pronunciation for stress, intonation, and the production of phonemes, as well as their grammar and vocabulary. Their interaction with Pi also served as a listening practice for them, where Pi would talk about something to them and then ask them comprehension questions. Learners could also ask Pi to switch accents if they were willing to expose themselves to different accents.

Pi-mediated activities were systematically designed to correspond with units 3-6 of the coursebook and thematically align with the American English File 3 curriculum. The instructor used Pi's analytics to monitor students' engagement regarding the time spent with Pi, topic mastery, and error patterns. This enabled the instructor to modify the lesson plan on a week-to-week basis to focus on key omissions or errors and provided a measure of monitoring students' individual progress in any given session.

At the end of the intervention, students from both groups completed a test of language achievement and the SOMS. The researcher also conducted semi-structured interviews with the participants of the experimental group to dig deeper into their perceptions regarding the practical applications of CHALL and its effects on their language achievement and sense of metavolvement.

2.4. Data Analysis

Paired-samples t-tests were used to examine within-group differences in language achievement and the sense of metavolvement. In addition, independent-samples t-tests were used to check differences in post-intervention language achievement between the experimental and the control group. The descriptive statistics of in-class and out-of-class metavolvement scores were then used to analyze the two groups of learners.

To analyze the interview data, a thematic analysis was conducted using Braun and Clarke's (2006) six-phase framework for thematic analysis, which followed these steps: 1) transcribing the interviews and familiarizing oneself with the transcripts, 2) generating initial codes, 3) searching for themes, 4) reviewing themes, 5) defining and naming themes, and 6) producing the report. For the purpose of ensuring the reliability of the analysis, a second coder coded 25% of the transcripts independently to establish an inter-rater reliability coefficient of $\kappa = 0.82$ (Armstrong et al., 1997). Irregularities were addressed through discussions, negotiations, modifications to the codebook, and outlines of the themes (Nowell et al., 2017). Finally, the primary researcher coded the transcripts, ensuring consistent coding.

3.Results

3.1. Quantitative Results

3.1.1. Validation of the Sense of Metavolvement Scale (SOMS)

Tests of reliability and CFA were undertaken to validate SOMS. The mean and standard deviation for SOMS and its subconstructs (in-class and out-of-class) are shown in Table 1. As can be seen, the mean values show that learners were engaged at quite high levels, particularly in the classroom context.

Table 3. *Descriptive Statistics for SOMS*

	Min	Max	Mean	SD
SOMS	13	60	47.10	7.73
In-class	6	30	24.18	4.13
Out-of-class	7	30	22.92	3.99

The normality of data distribution was checked by skewness and kurtosis values shown in Table 2. As can be seen, the values are well within the acceptable range of -2 to +2, indicating the normal distribution of the dataset.

Table 4. *Normality Tests for SOMS*

	Skewness	Kurtosis
SOMS	-.75	1.63
In-class	-.91	1.67

Out-of-class	-.45	.82

The reliability of SOMS was established through an analysis of Cronbach's alpha coefficients, which revealed sound internal consistency (see Table 3). It had an overall alpha coefficient of .90, with .82 for in-class and .84 for out-of-class metavolvement, indicating its reliability.

Table 5. *Reliability Estimates for SOMS*

	No. of Items	Cronbach's Alpha
SOMS	12	.90
In-class	6	.82
Out-of-class	6	.84

CFA was run to substantiate the construct validity of the SOMS scale. Harman's single-factor test was administered before CFA. The first factor accounted for only 49.18% of the total variance, confirming the multidimensionality of the model. Two subconstructs of in-class (6 items) and out-of-class (6 items) comprised the scale. Figure 2 illustrates the standardized factor loadings. No item was removed to enhance the model fit. The goodness-of-fit statistics are reported in Table 4.

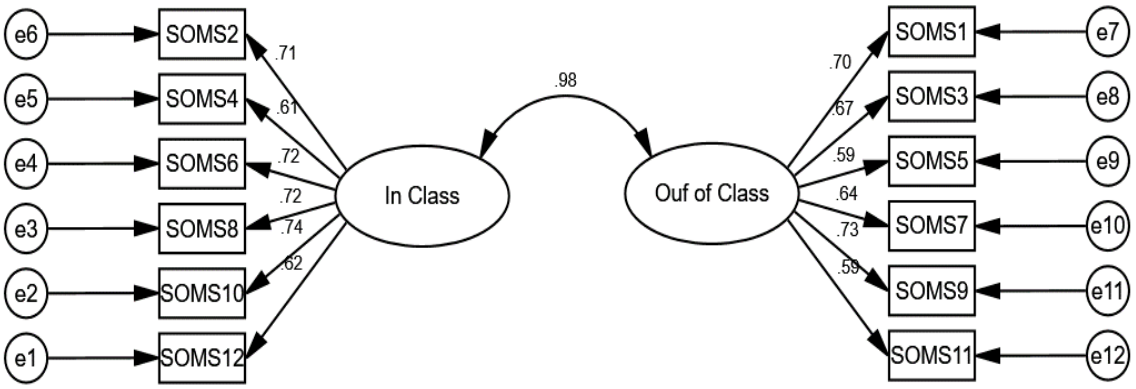

Figure 2. *Measurement Model for the SOMS Scale*

To see whether the model fit the data, the following goodness-of-fit indices were calculated using Amos: chi-square index divided by the degrees of freedom (χ^2/df), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Squared Error (SRMR). In this study, χ^2/df of less than 3 (Ullman, 2001), TLI and CFI of greater than .90, and RMSEA and SRMR of equal to or less than .08 (Browne & Cudeck, 1993) were considered indicators of an acceptable fit. As can be seen in Table 5, the model fits the data acceptably.

Table 6. *Goodness of Fit Indices for the Model*

Models	χ^2/df	Df	CFI	TLI	RMSEA	SRMR
CFA (Figure 1)	2.15	53	.94	.93	.07	.05

3.1.2. Effects of CHALL on Language Achievement

To examine the effects of CHALL on language achievement, paired-samples t-tests were conducted within each group to assess pre-test and post-test differences. Independent-samples t-tests were also used to compare post-test scores between the experimental and the control group. The results of the paired-samples t-test revealed a statistically significant improvement in the language achievement of the experimental group ($t(21) = -2.34$, $p = 0.03$) from the pre-test ($M = 67.21$, $SD = 7.89$) to the post-test ($M = 74.98$, $SD = 8.12$), suggesting that CHALL had a positive impact on learners' proficiency. In contrast, the control group showed no statistically significant improvement ($t(21) = 0.54$, $p = 0.60$) from pre-test ($M = 66.89$, $SD = 8.22$) to post-test ($M = 68.07$, $SD = 8.36$), indicating that conventional classroom instruction alone did not lead to measurable proficiency gains within the study period.

The results of the independent-samples t-test comparing post-test scores between the experimental and control groups showed no statistically significant difference ($t(42) = -0.72$, $p = 0.47$). While the CHALL group demonstrated notable within-group improvements, their performance did not significantly surpass that of the control group in post-test comparisons.

3.1.3. Effects of CHALL on Learners' Sense of Metavolvement

To determine whether CHALL significantly influenced learners' sense of metavolvement, pre-test and post-test SOMS scores were analyzed using paired-samples and independent-samples t-tests. The results of the paired-samples t-test showed that the experimental group had a significant increase in their overall metavolvement ($t(21) = -3.87$, $p = 0.001$), moving from a pre-test mean of 3.62 ($SD = 0.78$) to a post-test mean of 4.31 ($SD = 0.72$). The largest improvement was observed in in-class metavolvement ($t(21) = -3.16$, $p = 0.003$), whereas out-of-class metavolvement showed only a marginal increase ($t(21) = -1.84$, $p = 0.073$). In the control group, no statistically significant changes were observed in metavolvement ($t(21) = -0.92$, $p = 0.36$), indicating that conventional classroom instruction alone did not enhance learners' engagement levels.

The results of the independent-samples t-test comparing post-test scores between the experimental and control groups showed that the experimental group's post-test metavolvement scores were significantly higher than those of the control group ($t(42) = -2.88$, $p = 0.006$). This confirms that CHALL had a strong positive effect on learners' engagement and sense of ownership over their learning process.

3.2. Qualitative Results

3.2.1. Learner Perceptions of CHALL

The qualitative findings of semi-structured interviews with the participants in the experimental group provided some ideas about the advantages and limitations of using CHALL for language learning, as well as the effects of CHALL on engagement levels. The details are presented in Table 6.

Table 7. *Learners' Perceptions of CHALL*

Dimensions	Aspects	Themes	Definitions
Impact on language learning	Language skills	Reading	Ability to understand written texts
		Speaking	Ability to express ideas effectively in oral form
		Listening	Ability to comprehend spoken language
		Writing	Ability to express ideas effectively in written form
	Language components	Pronunciation	Accuracy in producing speech sounds
		Grammar	Understanding and correctly applying language rules and structures
		Vocabulary	The process of learning and retaining new words and phrases
	Cultural competence	Access to cultural information	Exposure to target culture knowledge
		Comparative cultural insights	Recognizing cultural nuances in language
		Cultural context in language use	Using language in a way that respects and aligns with cultural norms
User experience and engagement	Personalization and adaptability	Adaptation to the users' levels	Adjusting to learners' proficiency
		Customizable learning pace	Progressing at a preferred speed
		Tailored content suggestions	Recommending relevant materials
	Accessibility and convenience	24/7 availability	Continuous access to resources at any time
		Mobile learning	Learning on portable devices
		Flexible learning schedule	Learning at convenient times
	App features and functionality	Multi-modal learning	Incorporating various forms of input (text, audio, visual) in the learning process
		Voice customization	Option to adjust voice settings in CHALL for personalized listening experiences
		Conversation history	CHALL's feature of maintaining a record of past interactions for review and learning
	Motivation and confidence	Non-judgmental practice environment	A learning space free from criticism, encouraging experimentation with language
		Increased confidence in speaking	Greater self-assurance in oral communication resulting from the use of CHALL
Comparison with conventional methods	Advantages of CHALL	Immediate feedback	Instant responses and corrections from CHALL
		Relevant content	Current and relevant language content provided by CHALL
		Comprehensive skill development	Simultaneous improvement of multiple language skills
	Limitations and challenges	Accuracy concerns	Potential reliability issues with CHALL-provided information
		Limited effectiveness for beginners	Possible challenges for novice learners using CHALL
		Potential for overreliance	Risk of excessive dependence on CHALL
	Integration with conventional learning	Complementary use with conventional methods	Integration of CHALL with conventional teaching methods
		Enhanced classroom communication	Improved participation in conventional learning environments
		Preparation for real-life language use	Bridging classroom learning and practical language application

3.2.1.1. Impact on Language Learning

3.2.1.1.1. Language Skill Development

Regarding language skill development, the learners reported progress in reading, speaking, listening, and writing. The skill impacted the most in learners' view was reading. Some learners (14.2%) specified that CHALL improved their reading comprehension, as it tailored the reading text to their proficiency and provided them with context-based comprehension questions. One learner commented, "The chatbot can produce unique reading exercises for my level. That really helps me practice." This is indicative of the personalization ability of CHALL and its capacity to facilitate reading skill development by adapting reading passages to learners' proficiency levels.

The next skill impacted the most following the use of CHALL was speaking, with 11.6% of learners referring to their improved fluency following the use of CHALL. Many learners liked that the chatbot provided a space to practice speaking free of judgment. One learner shared, "I feel confident speaking with the chatbot because it never judges me, and I can practice with no fear of making mistakes." Another learner expressed, "I am able to experiment with new words and expressions without worrying about how I sound." Learners acknowledged that the immediate feedback on pronunciation assists them in improving their intonation.

A group of learners (9.5%) also indicated that CHALL helped them develop their listening comprehension, as there were many different speech rates, accents, and contexts. The learners appreciated the text-to-speech diversity and accents. One participant stated, "I understand different accents and faster speech much better thanks to the chatbot." A few learners, on the other hand, said that while CHALL exposed them to diverse accents, it did not fully mimic the complexity of real conversational speech, which dynamically uses multiple accents and registers.

Likewise, a total of 9.0% of the participants indicated that CHALL positively impacted their writing in terms of grammar, sentence structure, and organization. Participants also expressed the role of chatbot's suggestions in connecting ideas together. As one participant explained, "The chatbot helped me think about the structure of my sentences and gave me suggestions that made sense and made my writing clearer." Several learners, on the other hand, expressed that higher-order, more comprehensive feedback was needed for discourse-level organization and argumentation.

3.2.1.1.2. Language Components

The impact of CHALL on language components, including pronunciation, grammar, and vocabulary, was similarly positive. According to 10.9% of learners, feedback on pronunciation given by the chatbot helped them to correct their pronunciation errors and to refine their speaking skills. Several learners reported that the correct pronunciation of the chatbot was a better model for them to learn how to pronounce words accurately. One participant said, "The chatbot can model pronunciation and stress patterns in a way that's clear and easy to follow." However, some learners expressed that CHALL was not always able to provide feedback for other more complex elements, such as prosody and intonation patterns. Furthermore, some learners with stronger pronunciation skills felt that the pronunciation feedback from the chatbot

was much more rudimentary than feedback from a human teacher. It was perfectly suitable for learners with weaker pronunciation skills, but it did not handle the more nuanced features in pronunciation that stronger learners have to deal with.

In regard to grammar, 10.4% of learners said they found the instant feedback of the chatbot very useful. Learners liked how the chatbot provided contextual help to comprehend grammatical structures. As one participant noted, “The chatbot didn’t just correct me and say what was wrong but explained why I needed to remember the rule.” The rather instant feedback on grammar was an effective and low-pressure way for learners to improve accuracy.

As for vocabulary, 9.9% of learners felt CHALL was highly effective in expanding their vocabulary range. More specifically, learners appreciated that the chatbot presented new words within the context rather than just isolated forms. One learner said, “It taught me the new words in sentences that make sense and not just lists of vocabulary.” The contextualization of vocabulary aided the students in their comprehension of how any particular word functions within each context. Being provided with synonyms and example sentences also served to expand their overall vocabulary. Although contextualized learning has been shown to promote retention, some participants said a greater focus on collocation activities would better facilitate their vocabulary growth and improve their ability to apply words in natural ways in different contexts.

3.2.1.1.3. Cultural Competence

With respect to cultural competence, 7.1% of learners expressed that CHALL helped them to engage with various aspects of culture, such as customs, values, and historical contexts. Although this was a less prominent theme, it is still meaningful. In the current study, learners that interacted with different cultural content through the use of Pi had an enhanced awareness of cultural differences and were better able to interpret social cues as a result of their interaction. One participant stated, “I asked the chatbot about Canadian holidays and how people celebrate them. I felt like I got a better understanding of their values and why people act differently in conversing.” This represents the potential of chatbots in scaffolding intercultural learning in a way perhaps a coursebook or static content might not. On the flipside, many learners did not engage with this function, which may suggest that cultural content in CHALL is either too covert or not probed sufficiently. For this reason, it may be useful to include enhanced or even explicit cultural prompts or scenarios within the chatbot interactions, which may allow for deeper engagement and development of intercultural communicative competence.

3.2.1.2. User Experience and Engagement

A considerable amount of discussion in the interviews revolved around CHALL’s availability, with 13.2% of learners mentioning that its 24/7 availability offers them the flexibility to practice English anytime. A student stated, “Language learners can use the chatbot whenever and wherever they want.” Another major advantage of CHALL was personalization and adaptability, with 12.7% of learners appreciating CHALL’s flexibility in adjusting the content to their proficiency level. One student said, “I really find it interesting that it speaks to me based on my level.” In addition, a customizable learning pace, combined with tailored content

suggestions, was highly regarded, as it enabled learners to practice at their own pace and receive relevant material tailored to their needs. Despite the aforementioned advantages, the chatbot did not work for 6.5% of learners, as it sometimes failed to respond appropriately to their requests and misunderstood their intended meaning. Therefore, improved responsiveness of the chatbot and user interface can improve the overall user experience.

3.2.1.3. Motivation and Confidence

The lack of judgment in the feedback was a recurring theme for CHALL, with 11.3% of respondents stating that CHALL encouraged risk-taking in their use of language. One respondent explained, "Because it is artificial intelligence and has no judgment, it is really good," demonstrating how the chatbot reduced the fear normally present in language learning. Another participant stated, "I speak much more comfortably to my teacher in my English class." In addition, motivation and interest in language learning were also often emphasized, with 10.9% of respondents admitting that CHALL increased their motivation to learn the language. Some learners even reported that their interest extended beyond the boundaries of CHALL, with one respondent commenting, "I got a lot of motivation and even went on to start watching the series 'Friends'."

3.2.1.3. Comparison with Conventional Methods

A total of 10.4% of the respondents saw CHALL as more interactive, customized, and accessible compared to conventional methods. One key positive feature noted was finding out information in a timely manner. One participant stated, "It is much easier for me now to find an article or ask my questions about a problem compared to conventional methods." However, 6.8% of the participants also discovered the limitations of CHALL, especially for beginner students who might continue stumbling if they did not have structure from educators. Furthermore, there were participants that cautioned about a dependence on AI, mentioning that it might not allow for problem-solving skills to develop. One participant noted, "You might get used to it, and it might lead to solitude and loneliness."

3.2.1.5. Integration with Conventional Learning

Finally, the study participants unanimously agreed that while CHALL offers unique benefits, it could be most effective when used as a supplement and component of classroom instruction and not as a substitute. Learners maintained that the chatbot addressed temporal and spatial constraints of conventional learning, as one user noted, "Traditional methods are limited by time and location—CHALL lets me practice anytime, like reviewing grammar on my commute." This synergy also enhanced classroom engagement, with learners applying CHALL-driven practice to formal settings. One learner mentioned, "I used it to draft essays, then refined them with my teacher's feedback," implying the benefits of integrating CHALL with conventional learning.

4. Discussion

4.1. Validity and Reliability of the Sense of Metavolvement Scale

The SOMS, which was developed in the present study, was validated as a two-factor structure encompassing in-class and out-of-class metavolvement, which aligns well with previous

models that conceptualize engagement as a multidimensional construct including cognitive, emotional, and behavioral domains (Pishghadam et al., 2019b). The two-factor structure is also consistent with Fryer et al.'s (2020) study on AI tools, advocating for domain-specific measures to capture subtle interactions developed by learners. The findings also confirmed SOMS as a highly reliable scale.

4.2. The Effects of CHALL on Language Achievement

There were meaningful within-group differences in the experimental group's overall language achievement from pre- to post-intervention over the five-week period, which indicates proficiency gains following the use of CHALL. This is consistent with the findings of Kim et al. (2021), asserting that chatbots improve language proficiency through repetitive, low-anxiety practice. Previous studies that investigated the effects of CHALL on individual language skills and components have also confirmed that CHALL facilitates the development of various skills and components (Nagata et al., 2020; Qasem et al., 2023; Wu & Li, 2024; Yang et al., 2021; Yang et al., 2022; Zhang et al., 2023).

In line with previous research (Kim et al., 2021), this study also found that most learners appreciated CHALL because of its low-anxiety characteristic, which helped them reduce their speaking anxiety and in turn improved their overall language ability. This is also in accordance with Krashen's (1989) affective filter hypothesis, denoting the impact of psychological factors, such as stress, on language learning.

Despite meaningful within-group improvements in the experimental group, their post-test performance in language achievement was not significantly different from that of the control group. One possible reason for the lack of a significant difference between the two groups is that while CHALL was effective enough to improve language achievement in the experimental group, the positive impact was not major enough to create between-group differences, possibly due to the short duration of the intervention. Another reason is that the curriculum-aligned assessments (*American English File 3*) mainly emphasized content-specific material to be tested (e.g., vocabulary and reading comprehension) as opposed to pragmatic or intercultural competencies (Huang et al., 2023). In other words, the representation of language achievement was limited to classroom-aligned, regimented tasks and did not capture competencies outside the curriculum, which might have led to the lack of a significant difference between the two groups. Therefore, future studies should consider longitudinal designs on a larger sample size and use standardized assessments to measure proficiency gains.

It is also worth mentioning that this study focused on the impact of CHALL on overall language achievement and did not investigate how CHALL impacts basic language skills and components. Therefore, future studies can dig deeper into the impact of CHALL on various language skills, such as reading, listening, writing, and speaking, as well as grammar, vocabulary, and pronunciation.

4.3. CHALL's Impact on Learner Engagement and Metavolvement

Based on the findings, the experimental group demonstrated a statistically significant difference in their in-class metavolvement, which indicates CHALL's success in stimulating engagement in guided, teacher-led activities. In addition, the experimental group's post-test

metavolvement scores were significantly higher than those of the control group, confirming that CHALL had a strong positive effect on learners' sense of metavolvement. This finding aligns with the Interaction Hypothesis (Long, 1996), which asserts the role of interaction and meaning negotiation in language acquisition and learner engagement. This is also in line with the finding of Pishghadam et al. (2023) underscoring the positive correlation between the degree of support and concern given by the teacher and learners' success. The chatbot-based learning tasks (role-plays and grammar drills) afforded learners the necessary opportunities to negotiate meaning and receive corrective feedback. Furthermore, learners' remarks about the "non-judgmental environment" and "freedom to explore different language forms" suggest that CHALL lowered their anxiety about experimenting with language forms—a result consistent with Fryer et al.'s (2020) study on AI and psychological safety.

Regarding the out-of-class metavolvement, there was marginal improvement, which contradicts the findings of Wei's (2023) study reporting that learners sustained their engagement in self-regulated learning contexts autonomously. The poor finding in our study's out-of-class learning could possibly be attributed to the teacher-centered pedagogical style, where learners needed direction from the instructor to complete tasks assigned to them. The qualitative data also indicated that students utilized CHALL for "homework assigned by the teacher" and not on their own initiative, which results from the cultural norms encouraging structured guidance rather than learner autonomy. Culturally, the reliance on teacher-centric use of CHALL forms (i.e., homework assignments) represents pedagogies prevalent in Iran, which are heavily dependent on teachers with less emphasis on autonomy (Wei, 2023). This reliance would make it difficult to fully support CHALL in autonomous and self-regulated learning. Moreover, while participants' engagement with CHALL in structured ways improved their overall performance, it also dampened their chances for free use of CHALL on their own.

It is worth noting, however, that the study sample limits generalizability due to its size and homogeneity. Factors such as economic status (all participants were from a fee-paying academy), language skills (all intermediate), educational background, and age have possibly influenced CHALL's effectiveness but were not controlled or tested in the present study. Additionally, the relatively brief duration of the intervention (five weeks) did not allow us to explore the effects of long-term engagement in language. Therefore, future research should employ larger, heterogeneous samples to examine how socioeconomic status, age, and prior language skills mediate CHALL's impact. Further studies are also needed to examine the effects of CHALL on language skills and autonomous learning over extended periods of time.

4.4. Learners' Perceptions of CHALL

Regarding language skills and components, learners perceived CHALL to be beneficial for improving both. As for skills, they indicated that CHALL impacted their reading skills the most, followed by speaking, listening, and writing. As for components, they felt the greatest impact on their pronunciation, followed by grammar and vocabulary. Learners' perception regarding the impacts of CHALL on language skills and components perfectly aligns with the findings of previous studies. In line with previous research (Yang et al., 2022), learners also asserted that chatbots' ability to adapt reading tasks to their levels improved their reading skills, while the naturally varied accents and speech rates of chatbots enhanced their listening

comprehension. Furthermore, as previous studies have shown (Zhang et al., 2023), learners also recognized the role of chatbots in improving their writing skills by developing grammatical accuracy, coherence, syntactic variety, and lexical sophistication. In terms of grammar and vocabulary, consistent with previous studies (Qasem et al., 2023; Zhang et al., 2023), learners also expressed that CHALL expanded their grammar and vocabulary through corrective feedback and contextualized explanations.

On the other hand, as some previous studies highlighted chatbots' weaknesses in some areas, some learners also maintained the same view. More specifically, regarding pronunciation and listening, some learners expressed the chatbot was unable to mimic real conversational speech using multiple accents and registers dynamically, which was in agreement with previous studies referring to chatbots' poor representation of actual communication due to their limited variations in voice and real-life conversation features (Duong & Suppasetseree, 2024). As for writing, some learners expressed that feedback was mainly sentence-level, so it could not improve higher-order, discourse-level organization and argumentation, which aligns with previous research showing that chatbots often neglect higher-order writing aspects, such as argument development or coherence across paragraphs (Panda & Kaur, 2023). Regarding pronunciation, in line with previous research (Yang et al., 2022), learners also posited that chatbots have difficulty recognizing and correcting more complex errors related to prosody or intonation.

In addition, participants regarded CHALL as beneficial in lowering their anxiety and providing the opportunity to practice at their convenience. Most learners acknowledged that interaction with the chatbot was free of any judgments and full of encouragement to experience with the language. This is consistent with Krashen's (1989) affective filter hypothesis, which emphasizes the role of affective filters, such as anxiety, on language learning, and Fryer et al.'s (2020) finding that AI tools provide psychologically safe spaces for experimenting with the language. Learners also valued the chatbot's availability, flexibility, and personalization, as well as its positive role in enhancing their cultural understanding. Nonetheless, criticisms of CHALL's adherence to cultural norms, such as its inability to process Persian norms of politeness or idioms such as *khoda hafez*, demonstrate a significant gap in AI technology design, particularly in non-Western contexts (Patel et al., 2023). These limitations clearly indicate that human instructors have an indispensable role in fostering intercultural competence, specifically in countries like Iran where language and social norms are wholly intertwined. Therefore, the findings verify that CHALL can supplement conventional classroom instruction but not replace it, as Zhang et al. (2023) suggested in their meta-analysis. In other words, the findings suggest that CHALL is most effective as a hybrid learning model, used in conjunction with teacher instruction for skills practice along with higher-order competencies.

Based on the findings, AI systems would benefit from the development of feedback that is sensitive to emotion to support learners' emotional and cognitive processing (Yang et al., 2022). As technological advancements will allow for even more advanced AI models in the future, it would be interesting for future studies to evaluate increasingly sophisticated chatbots in productive contexts, particularly with regard to personalization and real-world context.

Future studies may also evaluate the value of multimodal AI interaction that can support enriched learner engagement and retention in learning experiences. Furthermore, future attempts could study CHALL's perceptions in a public institution, taking into account the cultural contexts in which it would be applicable to see if participants will engage with the platform over longer periods of time and in meaningful ways.

4.5. Pedagogical Implications

The current research provides some implications for pedagogy regarding the potential facilitative role of AI chatbots in language learning. Based on the findings, it may be wise to consider CHALL as just an additional educational tool to supplement classroom instruction, not to substitute it. For instance, while AI chatbots can assist learning at the level of pronunciation, grammar, and vocabulary immediately, at the level of discourse, learning requires human-mediated interaction and engagement. Furthermore, pragmatic competence and other learning-by-doing activities, which are at a higher level of cognition, including argumentation, also require human-mediated learning. As Zhang et al. (2023) explained, AI tools can be used successfully for controlled pedagogical tasks, but they cannot replace human instructors in other more complicated learning tasks. Educators should think about the possibility of implementing blended learning strategies in the classroom by using CHALL to facilitate the learning of specific skills, such as pronunciation and fluency, along with conventional classroom instructions for discourse-level skills. Such a balance can ensure that learners experience how effective an AI chatbot can be for their personal feedback and practice while simultaneously benefiting from an interactive and context-specific instruction by a teacher.

Conclusion

This study showed that Chatbot-Assisted Language Learning (CHALL) improved Iranian EFL learners' in-class metavolvement and overall language achievement in a structured and instructor-led setting. It also created significant improvements in the experimental group's sense of metavolvement compared to the control group. However, CHALL failed to exceed conventional instruction with respect to post-test analyses of language achievement and out-of-class metavolvement. This indicates that CHALL can complement conventional classroom instruction but cannot act as a distributor of knowledge on its own. Cultural shortcomings also highlight the importance of further investigation into developing AI tools that are sensitive to local cultures (Patel et al., 2023).

Given the methodological issues discussed (e.g., short intervention, homogeneous sampling), it is essential to conduct longitudinal investigations on CHALL with samples representative of the population overlooked by previous studies and to assess language achievement through standardized tests. Educators should encourage students to use CHALL for targeted skill practice (e.g., grammar, vocabulary) while using conventional human-mediated instructions for enumerative or inductive higher-order cognitive skills.

Acknowledgements

This work was supported by Ferdowsi University of Mashhad (FUM) under the grant number 60282.

References

Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology,* 31(3), 597–606.

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., ... & Pandey, A. K. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics,* 10(3), 325–341. https://doi.org/10.1007/s12369-018-0477-0

Bernstein, M. J., Sacco, D. F., Yong, S. G., Hugenberg, K., & Cook, E. (2010). Being "in" with the in-crowd: The effects of social exclusion and inclusion are enhanced by the perceived essentialism of ingroups and outgroups. *Personality and Social Psychology Bulletin,* 36(8), 999–1009. https://doi.org/10.1177/0146167210376059

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge University Press.

Duong, T. & Suppasetseree, S. (2024). the effects of an artificial intelligence voice chatbot on improving vietnamese undergraduate students' English speaking skills. International Journal of Learning Teaching and Educational Research, 23(3), 293-321. https://doi.org/10.26803/ijlter.23.3.1

Fryer, L., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology,* 24(2), 8–22.

Gholami, M. J. and Al Abdwani, T. (2024). The Rise of Thinking Machines: A Review of Artificial Intelligence in Contemporary Communication. Journal of Business, Communication & Technology, 3(1), 29-43. https://doi.org/10.56632/bct.2024.3103

Golonka, E. M., Bowles, A. R., Frank, V., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning,* 27(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning,* 38(1), 237–257. https://doi.org/10.1111/jcal.12505

Kim, H. S., Kim, N. Y., & Cha, Y. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance? *Journal of Asia TEFL,* 18(1), 161–178.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal,* 73(4), 440–463. https://doi.org/10.1111/j.1540-4781.1989.tb05325.x

Latham-Koenig, C., Oxenden, C., & Lambert, J. (2020). *American English File 3* (2nd ed.). Oxford University Press.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition: Vol. 2. Second language acquisition* (pp. 413–468). Academic Press.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265–302. https://doi.org/10.1017/S0272263109990520

Nagata, R., Hashiguchi, T., & Sadoun, D. (2020). Is the simplest chatbot effective in English writing learning assistance? In M. D. Lytras, P. Ordoñez de Pablos, & M. Helfert (Eds.), *Computational linguistics: 16th International Conference of the Pacific Association for Computational Linguistics, PACLING 2019, Hanoi, Vietnam, October 11–13, 2019, Revised Selected Papers* (pp. 245–256). Springer Singapore.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1–13.

Panda, S., & Kaur, N. (2023). Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers. *Library Hi Tech News*, 40(3), 22–25.

Patel, R., Al-Hassan, S., & Kim, Y. (2023). Cultural responsiveness in AI language tools: A critical review. *ReCALL*, 35(2), 210–225. https://doi.org/10.1017/S0958344023000079

Pishghadam, R. (2016). Introducing emotioncy tension as a potential source of identity crisis. *Paper presented at the 16th Interdisciplinary Conference on Cultural Identity and Philosophy of the Self*, London, England.

Pishghadam, R., & Shayesteh, S. (2017). Emo-sensory expression at the crossroads of emotion, sense, and language: A case of color-emotion associations. *International Journal of Society, Culture, & Language (IJSCL*), 5(2), 15–25.

Pishghadam, R., Al Abdwani, T., Jajarmi, H., & Shayesteh, S. (2023a). Enhancing general and language aptitude tests by incorporating cultural and emo-sensory constructs. *International Journal of Society, Culture and Language*, 11(3), 1–12. https://doi.org/10.22034/ijscl.2023.704891

Pishghadam, R., Baghaei, P., & Seyednozadi, Z. (2017). Introducing emotioncy as a potential source of test bias: A mixed Rasch modeling study. *International Journal of Testing*, 17(2), 127–140. https://doi.org/10.1080/15305058.2016.1183208

Pishghadam, R., Ebrahimi, S., & Tabatabaiean, M. (2019a). *A Novel Approach to Psychology of Language Education*. FUM Press.

Pishghadam, R., Ebrahimi, S., Rajabi Esterabadi, A., & Parsae, A. (2023b). Emotions and success in education: From apathy to transpathy. *Journal of Cognition, Emotion & Education*, 1(1), 1–16. https://doi.org/10.22034/cee.2023.172495

Pishghadam, R., Shakeebaee, G., & Rahmani, S. (2019b). Sensory capital in education: The missing piece? *Italian Journal of Sociology of Education*, 11, 1–30. https://doi.org/10.14658/pupj-ijse-2019-3-15

Qasem, F., Ghaleb, M., Mahdi, H. S., Al Khateeb, A., & Al Fadda, H. (2023). Dialog chatbot as an interactive online tool in enhancing ESP vocabulary learning. *Saudi Journal of Language Studies*, 3(2), 76–86.

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology*, 14(2), 95–110.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp. 653–771). Allyn & Bacon.

Viberg, O., & Grönlund, Å. (2013). Cross-cultural analysis of users' attitudes toward the use of mobile devices in second and foreign language learning in higher education: A case from Sweden and China. *Computers & Education,* 69, 169–180. https://doi.org/10.1016/j.compedu.2013.07.014

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching,* 31(2), 57–71. https://doi.org/10.1017/S0261444800012970

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. Frontiers in Psychology, 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wu, X., & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *The Asia-Pacific Education Researcher*. 33(3), 1-15. https://doi.org/10.1007/s40299-024-00853-2

Yang, H., & Kim, H. (2021). Development and application of AI chatbot for cabin crews. *Korea Journal of English Language and Linguistics,* 21(0), 1085–1104. http://journal.kasell.or.kr

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL,* 34(3), 327–343. https://doi.org/10.1017/S095834402200007X

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers & Education: Artificial Intelligence,* 4, Article 100134. https://doi.org/10.1016/j.caeai.2023.100134

Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: Pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*, 1–29. https://doi.org/10.1080/10494820.2023.2207430