Can AI chatbots effectively improve EFL learners' learning effects?—A meta-analysis of empirical research from 2022–2024

Zhaoyang Liu, Wenlan Zhang & Panpan Yang

Published online: 08 Feb 2025.

Submit your article to this journal

Article views: 1512

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Can AI chatbots effectively improve EFL learners' learning effects?—A meta-analysis of empirical research from 2022–2024

Zhaoyang Liu[a] (iD), Wenlan Zhang[a] (iD) and Panpan Yang[a,b] (iD)

[a]Faculty of Education, Shaanxi Normal University, Xi'an, China; [b]Faculty of Teacher Education, Hainan Normal University, Haikou, China

ABSTRACT

Technology development has brought transformative changes to the field of education, with AI chatbots injecting new vitality into language learning. To clarify the effectiveness of AI-powered chatbots in enhancing learners' English language acquisition, this study systematically employed meta-analysis to review 25 peer-reviewed studies published between 2022 and 2024. The results indicate that AI chatbots significantly enhance learners' academic performance, positively influencing their language skills as well as their psychological and emotional well-being. Key moderating factors include intervention duration, chatbot role, and chatbot form. Longer intervention periods correlated positively with better learning effects; AI chatbots as learning companions were more effective; and AI chatbots with mechanical appearances had a greater impact on learning. However, educational levels and teacher roles did not significantly affect the moderating effects. This paper offers practical recommendations for integrating AI chatbots into EFL learning and suggests future research directions, providing valuable evidence and insights into their effectiveness in EFL education.

1. Introduction

A Chatbot is a human-computer dialogue interaction system enabling users to communicate with computer programs (Kwon et al., 2023; Lee et al., 2020). In 1956, Joseph Weizenbaum developed ELIZA, the world's first chatbot, which employed keyword-matching technology for text

---

CONTACT Wenlan Zhang ✉ wenlan19@163.com 🖃 Faculty of Education, Shaanxi Normal University, Xi'an, China

input and output (Han & Lee, 2022; Weizenbaum, 1966). Early chatbots were limited to script question-and-answer interactions and lacked natural interactivity. However, With the rapid development of artificial intelligence technologies, such as natural language processing (NLP), machine learning (Jeon, 2024), speech recognition (Yuan, 2023), speech synthesis (Muniandy & Selvanathan, 2024), and deep learning (Ghafouri, 2024), the performance of chatbots has significantly improved (Chen et al., 2017). They can now be better able to perform tasks, execute commands, answer questions, provide detailed information (Kim et al., 2022), and deliver personalized services to learners (Hwang et al., 2022; Wang et al., 2022). In 2022, OpenAI introduced ChatGPT, an AI chatbots based on large language models (OpenAI, 2023), which generates human-like, high-quality responses based on user prompts. The rise of generative artificial intelligence (AIGC) has brought new opportunities to technology-assisted language learning, driving the rapid development of AI-based chatbots. Consequently, scholarly interest in the impact of AI chatbots on language learning has grown (Lee et al., 2023). Despite the proliferation of studies, findings remain inconsistent, and there is a lack of systematic reviews clarifying their effectiveness in language education. Therefore, this study employs a meta-analysis to systematically examine the impact of AI chatbots on the learning effects of English as a Foreign Language (EFL) learners, focusing on both language skills and psychological and emotional aspects. The study evaluates the effectiveness of AI chatbots in EFL learning, exploring potential moderating factors, and aims to provide practical guidance for educators and related professionals on leveraging AI technology to enhance foreign language education.

2. Literature review

2.1. AI chatbots in EFL learning

English as a Foreign Language (EFL) involves acquiring knowledge and skills related to foreign languages, with chatbots considered valuable tools for language learning (Lin & Mubarok, 2021; Yang et al., 2022). Due to limited classroom time and resources, teachers often find it challenging to provide timely and targeted feedback to all learners (Luo, 2016; Wang et al., 2022). Chatbots offer convenient opportunities for non-native learners of foreign languages (Jeon, 2024), allowing them to engage without time or geographical constraints (Chen et al., 2020; Fryer & Carpenter, 2006; Kwon et al., 2023; Rad et al., 2023). Learners can converse with chatbots in simulated scenarios free from peer pressure, providing an efficient, diverse, and personalized learning experience (Moussalli & Cardoso, 2020).

Research indicated that chatbots can significantly enhance learners' performance in listening, speaking (Moussalli & Cardoso, 2020), vocabulary (Görgen et al., 2020), and writing (Guo et al., 2022; Kwon et al., 2023; Rad et al., 2023), while also stimulating learning interest (Chen et al., 2020; Guo et al., 2022), enhancing learning motivation (Polyzi & Moussiades, 2023), foreign language enjoyment (Wang et al., 2022), and self-efficacy (Ryong et al., 2023; Zhang et al., 2023). AI chatbots provide a human-like, autonomous learning experience (Rad et al., 2023), effectively alleviating learners' anxiety (Hwang et al., 2022; Jeon, 2024; Wang et al., 2022; Zhang & Han, 2021). However, some studies have identified several drawbacks, including difficulties in understanding learners' intentions (Jeon, 2024), an inability to maintain continuous dialogue on the same topic (Fryer et al., 2019), lack of emotional engagement, concerns about data privacy, and issues with inaccurate or insufficient responses (Alemdag, 2023). These drawbacks may negatively impact learners, indicating that the influence of AI chatbots on language learning is more complex than initially assumed.

Existing research suggests that the effectiveness of AI chatbots in second language learning is primarily reflected in language acquisition, as well as psychology and emotional states (Rad et al., 2023; Wang et al., 2022). Language acquisition is closely related to the balanced development of listening, speaking, reading, and writing skills (Chen et al., 2020; Newton et al., 2018; Wang et al., 2022). Oral practice engages multiple sensory modalities in processing information, and sustained practice can boost learners' confidence, interest (Guo et al., 2022; Polyzi & Moussiades, 2023), language sensitivity, comprehension ability (Lee et al., 2023), reading ability (Görgen et al., 2020), and oral expression skills (Ye et al., 2022). Most learners enjoy interacting with AI chatbots for oral practice, language assessment (Zou et al., 2023), and pronunciation correction through repeated interactions (Fryer et al., 2019), which significantly enhances their speaking skills (Lin & Mubarok, 2021).

Vocabulary is a crucial language component that aids in understanding and acquiring a foreign language (Jeon, 2023; Polyzi & Moussiades, 2023). Research shows that AI chatbots can significantly enhance learners' vocabulary acquisition (Jeon, 2023) and interests (Polyzi & Moussiades, 2023). Yang et al. (2022) designed and developed a voice AI chatbots Ellie, and found that EFL learners from Korea preferred using common words, with interaction frequencies far exceeding those in traditional classrooms.

AI chatbots also offer meaningful guidance for EFL writing (Guo et al., 2022; Kwon et al., 2023; Rad et al., 2023). For instance, Kwon et al. (2023) found significant improvements in primary school students' second language writing after using 'Dialogflow'. Through mixed research,

Rad et al. (2023) discovered that AI chatbots (Wordtune) enhanced college students' engagement, feedback literacy, and writing performance in second language learning. However, Escalante et al. (2023) compared the effects of human tutor feedback and ChatGPT 4.0 feedback on learners' writing performance, finding no significant difference in scores between the two groups. Although AI chatbots can significantly save teachers' time, they do not provide better assistance to learners.

AI chatbots can evoke different emotional experiences in language learners (Ghafouri, 2024). They generate new learning materials based on learners' commands (Lee et al., 2023), aligning with their interests and improving engagement (Abdelhalim, 2024; Lee et al., 2023) and learning efficiency (Abdelhalim, 2024). Lee et al. (2023) used a control experiment to compare the impact of generative AI on primary school students' English reading learning. They found that AIGC technology improved learners' enjoyment and reading interest. Moreover, research also shows that learners enjoy using AI chatbots (Yuan, 2023), find them helpful in improving their language skills (Fryer & Carpenter, 2006), and feel more comfortable interacting with chatbots (Kwon et al., 2023; Tai & Chen, 2023; Wang et al., 2022). Human peers may exert pressure on learners (Wang et al., 2022), causing fear of mistakes or underperforming (Guo et al., 2022), whereas interacting with chatbots alleviates this pressure. Research indicates a positive correlation between second language enjoyment and learning effects (Wang et al., 2022). However, Tai and Chen (2023) found that learners' self-efficacy in using AI chatbots to assist EFL essay writing ($M = 130.00$, SD = 32.42) was slightly lower than before training ($M = 130.86$, SD = 25.06). Fryer et al. (2019) compared AI chatbots' performance in different tasks and learners' interest, finding that learners' interest in AI chatbots gradually declined over time, suggesting that the novelty effect significantly influences learners' use of AI chatbots.

In summary, despite current technological limitations, the potential of AI chatbots in EFL learning is undeniable. AI chatbots offer distinct advantages in enhancing learners' language skills and psychological well-being. However, existing research results remain inconsistent, and further validation of AI chatbots' effectiveness in language learning is necessary.

2.2. Prior Mata-Analyses of AI chatbots in EFL

A review of recent meta-analyses on the impact of AI chatbots on learners' academic outcomes reveals that most studies consider variables such as student grade level, intervention duration, interaction type, task type, and technology type as moderators (Alemdag, 2023; Wu & Li, 2024; Wu

& Yu, 2024). Alemdag (2023) analyzed 28 studies published before 2022, examining six moderating variables: educational level, subject area, control group teaching type, experiment duration, chatbot type, and chatbot task. The results revealed that chatbots significantly enhance learning effects.

Wu and Yu (2024) investigated the effects of AI chatbots on learners' academic outcomes, motivation, interest, self-efficacy, anxiety, and learning perceptions, analyzing the moderating effects of educational level and intervention duration. Their results indicated that AI chatbots significantly impact learners' academic performance. However, their study covered various disciplines such as computer science, medicine, linguistics, and psychology. In contrast, this study exclusively focuses on the EFL field, examining the impact of AI chatbots on EFL learners' learning effects, thus providing a more targeted analysis. Additionally, Wu and Yu (2024) analyzed papers published between 2019 and 2022, while this study includes literature from 2022 to 2024, incorporating the most recent research on AI chatbots in language learning, making the findings more timely and relevant.

Wu and Li (2024) conducted a meta-analysis on language learning, analyzing 29 studies published between 2004 and 2023. They examined the impact of robot-assisted language learning (RALL) on EFL learners' language skills, including listening, speaking, reading, writing, vocabulary, and overall language proficiency. The results showed that RALL improves learners' overall language ability. However, their study focused solely on the impact of AI chatbots on EFL learners' language skills, neglecting the importance of psychological and emotional factors such as motivation, self-efficacy, and learning interest, indicating certain limitations. Furthermore, earlier AI chatbots had limited intelligence, whereas advancements in AI technology, especially generative AI, have significantly enhanced AI chatbots' intelligence, potentially affecting their impact on learners' learning effects. In summary, while research has explored the influence of chatbots on academic performance, few meta-analyses have focused specifically on language learning. These meta-analyses have not thoroughly analyzed the diverse impact of AI chatbots on students' language skills and psychological conditions. The rapid development of generative AI has endowed AI chatbots with better content understanding and personalized feedback capabilities, bringing enormous potential to the language learning field. However, the extent to which AI chatbots affect EFL learners' academic outcomes remains unclear. Therefore, this study employs a quantitative synthesis to systematically review 25 studies published in core databases between 2022 and 2024, aiming to answer the following research questions: (1) Can AI chatbots improve EFL learners' learning effects? (2) How do different moderating factors influence EFL learners' learning effects?

3. Research design

This study follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure transparency and thoroughness in the systematic review process, which includes conducting a literature search, screening, assessing eligibility, selecting studies, extracting data, and coding (Page et al., 2021).

3.1. Data sources

This study utilized databases such as Web of Science, EBSCO, Wiley, ERIC, and ProQuest as data sources. The search query employed was: ('Artificial intelligence robot' OR 'AI chatbots' OR 'Chatbot' OR 'AI-assisted learning' OR 'AI-powered chatbots' OR 'ChatGPT' OR 'Generative Artificial Intelligence' OR 'generative AI') AND ('AI-assisted language learning' OR 'language learning' OR 'second language') AND ('Learning outcome' OR 'learning achievement' OR 'achievement' OR 'outcome' OR 'learning performance' OR 'learning effects' OR 'academic performance' OR 'Learning Effect'). To ensure comprehensive coverage, Google Scholar was also used to supplement the search, along with backward and forward citation tracking of the relevant literature and reference lists. The search was conducted up to April 2024.

3.2. Data filtering criteria

During the preliminary literature search, 3502 documents were identified after duplicates were removed. Relevant literature was selected based on the following criteria: (1) Articles must have undergone peer review and belonged to the field of education; (2) The research must focus on AI chatbots-assisted EFL learning; (3) The study subjects must be English learners at K-12 and higher education levels from non-English-speaking countries or regions with a high population of immigrants; (4) The research must employ experimental or quasi-experimental methods; (5) The literature must contain complete data necessary for meta-analysis, including sample size (N), mean (M), standard deviation (SD), or values such as Cohen's d or Hedges' g to represent standardized mean differences (SMD). After reviewing titles and abstracts, 2850 articles were retained. Subsequently, full-text reading were conducted, and articles that did not meet the specified criteria were excluded. Finally, 25 articles were included in the current systematic review. The detailed literature selection process is illustrated in Figure 1.
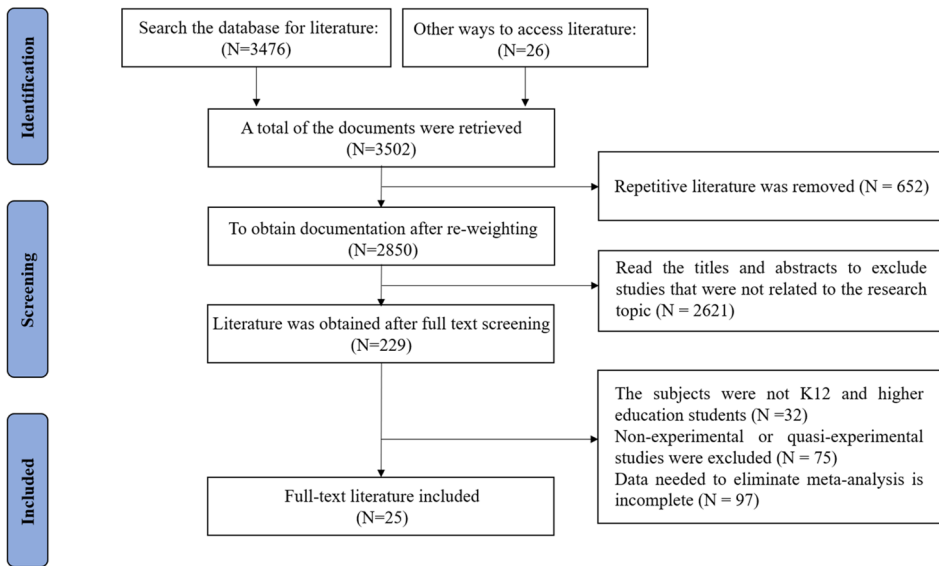
Figure 1. Literature screening process.

3.3. Document coding

High-quality data extraction and coding are critical for ensuring the reliability of results (Cooper et al., 2019). Drawing from existing meta-analyses, relevant research literature (Huang et al., 2022; Jeon, 2024), and the 25 empirical studies included in this research, two researchers consolidated synonymous, near-synonymous, or hierarchical concepts to develop the coding framework (Table 1). The coding framework comprises two parts: (1) Basic literature information, such as article number, title, author, publication date, country or region of publication, and foreign language proficiency. (2) Feature value coding, including intervention duration, education level, teacher role, robot role, robot appearance, sample size, and effect size calculation.

Intervention duration is categorized into short-term (≤ 1 week), medium-term (more than 1 week but less than 4 wk), and long-term (more than 4 wk but within one semester) (Randall, 2020). Wu and Yu (2024) suggest dividing education levels into primary, secondary and tertiary. Based on the research by Wu and Li (2024), the teacher's role is classified into three categories: teacher-led instruction, teacher-assisted AI chatbots instruction, and no teacher involvement. According to Randall (2020), the role of the chatbot is categorized as AI chatbots as teacher (the chatbot independently undertakes teaching tasks and guides student learning), AI chatbots as teaching assistant (the chatbot serves as a supplementary tool), and AI chatbots as learning companion (the chatbot interacts with students as a peer or learning partner) (Chang et al., 2010). Chatbot appearance is

Table 1. Coding framework.

Dimension	Category	Description	References
Intervention Durations	Short	≤1 week	(Randall, 2020)
	Intermediate	>1 week, ≤4 wk	
	Long	>4 wk, ≤one semester	
Educational Levels	Primary	Preschool or primary school students	(Wu & Yu, 2024)
	Secondary	Junior high school or senior high school students	
	Tertiary	College students	
Types of Teacher Role	Tutor	Teacher-led EFL lessons	(Wu & Li, 2024)
	Assistant	Teachers assisting Chatbots in teaching language	
	No teachers	No teachers available in EFL	
Types of Chatbot Role	Assistant	Chatbots acting as teacher assistants	(Randall, 2020)
	Peer	Chatbots acting as student peers	
	Teacher	Robot-led EFL lessons	
Types of Chatbot Form	Anthropomorphic	Human-like torso or facial features (e.g. NAO)	(Randall, 2020)
	Zoomorphic	Animal-like or caricature-like features (e.g. iCat)	
	Mechanomorphic	Machine-like qualities (e.g. ChatGPT)	
Language Skill	Listening (L1)	Listening development	(Wu & Li, 2024)
	Speaking (L2)	Speaking skill	
	Reading (L3)	Reading skill	
	Writing (L4)	Writing skill	
	Vocabulary (L5)	Vocabulary skill	
	Language in General (L6)	General language skill	
Psychological/ Emotion	Learning Motivation (E1)	Motivation to initiate and sustain learners' learning behavior	(Wu & Yu, 2024) (Jin & Zhang, 2021)
	Learning Interest (E2)	Affective disposition of learners towards engaging in learning activities	
	Perceived Value of Learning (E3)	Perceived value of learning activities by learners	
	Learning Self-Efficacy (E4)	Confidence in learners' abilities	
	Foreign Language Enjoyment (E5)	Enjoyment is one such positive emotion. Thus far, researchers have investigated its underlying dimensions)	

divided into three types: anthropomorphic, zoomorphic, and mechanomorphic. Wu and Li (2024) examined students' language abilities in six areas: listening, speaking, reading, writing, vocabulary, and overall language proficiency. The majority of existing studies evaluate these skills using tests, performance metrics, self-reports, and so on. Building on Jin and Zhang (2021), this study also explores the psychological and emotional impacts of AI chatbots use, examining factors like learning motivation, learning interest, value perception, self-efficacy, and foreign language enjoyment. Most studies assess these aspects through validated instruments focusing on variables such as learning motivation, self-efficacy, anxiety, engagement, and foreign language enjoyment. However, some studies have also utilized multimodal data to

analyze the psychological and emotional impacts of AI chatbots use, such as facial expressions and interaction data, which provide additional insights into learners' emotional responses.

The two researchers independently coded the data. Any discrepancies were discussed and collaboratively resolved, yielding a total of 70 effect sizes and 4,179 samples. The Cohen's Kappa coefficient for the coding was 0.91 (>0.7), indicating high reliability in the coding results. Specific coding details are the presented in Table 2.

3.4. Data analysis

This study employed CMA 3.0 (Comprehensive Meta-Analysis) for data analysis. Heterogeneity testing was conducted to evaluate the consistency of the included studies, using both the Q-test and I^2-test. If the p value of the Q-test exceeds 0.05, the studies are considered homogeneous, and a fixed-effects model is applied. Conversely, if $p < 0.05$, the studies are deemed heterogeneous, and a random-effects model is used. Generally, I^2 values of 25%, 50%, and 75% correspond to low, moderate, and high heterogeneity, respectively (Hedges et al., 1987). Since the I^2-test can sometimes provide more precise results, it is often used to complement the Q-test. In this study, heterogeneity testing yielded $p < 0.05$ and $I^2 = 92.86$, indicating high heterogeneity, and thus, a random-effects model was selected. The results of the heterogeneity testing are detailed in Table 3.

3.5. Publish offset test

Publication Bias arises when the included literature does not adequately represent the broader research landscape, potentially leading to an overestimation of the overall effect size. The Funnel Plot provides a visual representation of this bias, where greater asymmetry indicates a higher degree of bias. Egger's regression test offers a quantitative measure of publication bias (Wu & Yu, 2024). In this study, both methods were utilized to assess publication bias.

As shown in Figure 2, the Funnel Plot exhibits a symmetric inverted funnel shape, with data points evenly distributed in the upper-middle region. The Egger's test produced a p value of 0.594 (>0.05), indicating a low probability of publication bias. These results confirm the robustness and reliability of the study's findings.

4. Results

This study analyzed the impact of AI chatbots on learners' EFL learning effects, comparing the effects of moderator variables such as intervention

Table 2. Document coding list.

No.	Author(Year)	Country/Region	Number of Samples EG	Number of Samples CG	Intervention Durations	Educational Levels	Types of Teacher Role	Types of Robot Role	Control Group Teaching Methods	English Proficiency	Learning Effects
1	(Kim, 2022)	Korea	119	119	Long	Tertiary	Assistant	Teacher	Mechanomorphic	–	L6
2	(Chien et al., 2022)	Taiwan, China	37	36	Intermediate	Secondary	Tutor	Assistant	Mechanomorphic	–	L1, L2, E1
3	(Hwang et al., 2022)	Taiwan, China	23	20	Long	Primary	No	Teacher	Mechanomorphic	–	L2 L6
4	(Belda-Medina & Calvo-Ferrer, 2022) 1	Spain	115	115	Intermediate	Tertiary	No	Assistant	Anthropomorphic	B2-C1	E2
5	(Belda-Medina & Calvo-Ferrer, 2022) 1	Spain	61	61	Intermediate	Tertiary	No	Assistant	Anthropomorphic	B2-C1	E2
6	(Belda-Medina & Calvo-Ferrer, 2022) 2	Spain	115	115	Intermediate	Tertiary	No	Assistant	Zoomorphic	B2-C1	E2
7	(Belda-Medina & Calvo-Ferrer, 2022) 2	Spain	61	61	Intermediate	Tertiary	No	Assistant	Zoomorphic	B2-C1	E2
8	(Yang et al., 2022)	Korea	177	177	Short	Primary	Tutor	Peer	Anthropomorphic	A1-A2	E1, E2, E3, E5
9	(Wang et al., 2022)	China	327	327	Long	Primary	Tutor	Assistant	Anthropomorphic	–	L6, E5
10	(Ye et al., 2022)	China	28	28	Long	Secondary	Tutor	Peer	Zoomorphic	–	L2, L6
11	(Song & Song, 2023)	China	25	25	Long	Tertiary	Tutor	Assistant	Mechanomorphic	–	L4, L5, L6, E1
12	(Zou et al., 2023)	China	35	35	Long	Tertiary	Assistant	Assistant	Mechanomorphic	–	L2, L6
14	(Kwon et al., 2023)	Korea	37	38	Long	Primary	Tutor	Assistant	Mechanomorphic	–	L4
15	(Rad et al., 2023)	Iran	23	23	Long	Secondary	Tutor	Assistant	Mechanomorphic	–	L4, E1
16	(Hong et al., 2023)	Taiwan, China	153	153	–	Secondary	Assistant	Assistant	Mechanomorphic	–	L6, E3, E5
17	(Yuan, 2023)	Britain	39	35	Long	Primary	Tutor	Assistant	Anthropomorphic	–	L2, E1
18	(Lee et al., 2023)	Korea	60	61	Short	Primary	Tutor	Assistant	Mechanomorphic	–	L3, E2, E5
19	(Polyzi & Moussiades, 2023)1	Greece	10	10	Intermediate	Primary	Tutor	Assistant	Mechanomorphic	–	L5, E2, E3, E5
20	(Polyzi & Moussiades, 2023)2	Greece	10	10	Intermediate	Primary	Tutor	Assistant	Mechanomorphic	–	L5
21	(Shaikh et al., 2023)	Norway	10	10	Short	Tertiary	Assistant	Teacher	Mechanomorphic	–	L1, L2, L3, L4
22	(Jeon, 2023) 1a	Korea	18	17	Short	Primary	Tutor	Assistant	Mechanomorphic	–	L5
23	(Jeon, 2023) 1b	Korea	18	17	Short	Primary	Tutor	Assistant	Mechanomorphic	–	L6
24	(Jeon, 2023) 2a	Korea	18	17	Short	Primary	Tutor	Assistant	Mechanomorphic	–	L5
25	(Jeon, 2023) 2b	Korea	18	17	Short	Primary	Tutor	Assistant	Mechanomorphic	–	L6
26	(Escalante et al., 2023)	America	23	25	Long	Tertiary	Tutor	Teacher	Mechanomorphic	B1	L4

(Continued)

Table 2. Continued.

No.	Author(Year)	Country/Region	Number of Samples		Intervention Durations	Educational Levels	Types of Teacher Role	Types of Robot Role	Control Group Teaching Methods	English Proficiency	Learning Effects
			EG	CG							
27	(Zhang et al., 2023)	China	15	15	Long	Tertiary	No	Teacher	Mechanomorphic	–	L4, E4
28	(Tai & Chen, 2023)	China	112	112	Intermediate	Secondary	No	Peer	Mechanomorphic	–	E2, E4
29	(Ryong et al., 2023)1	Korea	22	24	Short	Tertiary	No	Teacher	Mechanomorphic	–	E2, E4, E5
30	(Ryong et al., 2023)2	Korea	33	24	Short	Tertiary	No	Teacher	Mechanomorphic	–	E2, E4, E5
31	(Lee et al., 2024)1	Korea	80	80	Long	Tertiary	No	Assistant	Mechanomorphic	–	L4
32	(Lee et al., 2024)2	Korea	80	80	Long	Tertiary	No	Assistant	Mechanomorphic	–	L4
33	(Ghafouri, 2024)	Iran	15	15	Long	Secondary	Tutor	Assistant	Mechanomorphic	B1	E1, E4
34	(Muniandy & Selvanathan, 2024)	Malaysia	40	40	Long	Tertiary	No	Teacher	Mechanomorphic	B1/C1	E2, E3, E4
35	(Abdelhalim, 2024)	Saudi Arabia	12	12	Long	Tertiary	No	Assistant	Mechanomorphic	B2	L6, E2, E3

Notes: CG, control group; EG, experimental group.

Table 3. Heterogeneity Test.

Effect model	Hedges' g	n	Standard error	95% (CI)		Heterogeneity test result			
				Lower	Upper	Q	df	I²	P
Fixed	0.66	70	0.02	0.61	0.72	962.97	69	92.86	<0.001
Random	0.97	70	0.11	0.76	1.18				

Notes: n, number of effect sizes; CI, confidence interval; Q, Cochran's Q statistic; df, degrees of freedom; I², heterogeneity percentage; P, p value.
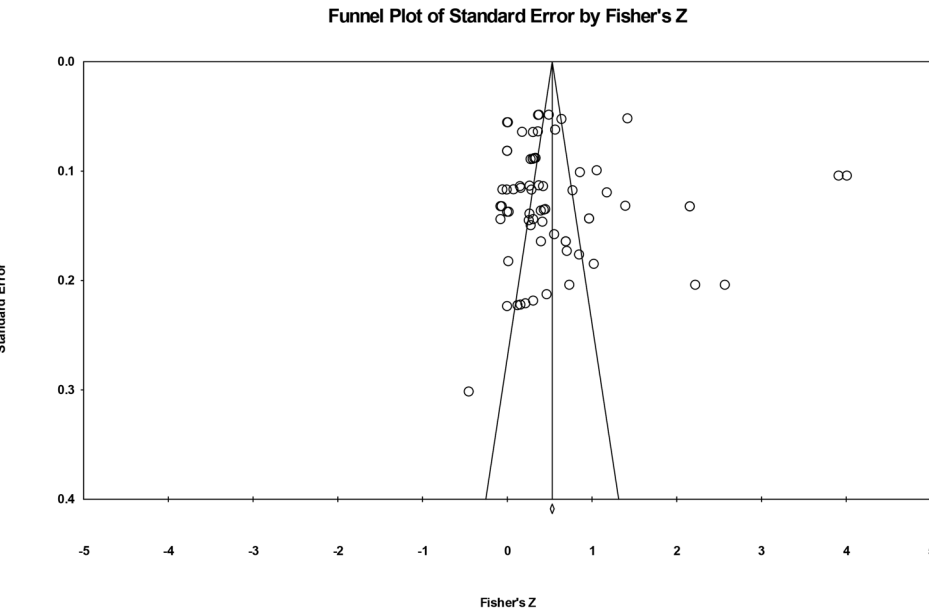*p < .05; **p < .01; ***p < .001.

Figure 2. Funnel diagram.

duration, educational level, teacher role, chatbot role, and chatbot form on EFL learning. The standardized mean difference was measured using Hedge's g, with values of 0.2, 0.5, and 0.8 representing small, moderate, and large effects, respectively (Wu & Yu, 2024).

4.1. AI chatbots have a significantly positive impact on learners' EFL learning effects

AI chatbots demonstrated a moderate to high positive effect on learners' overall EFL language skills ($g = .72 > 0.5$), with significant differences across groups ($p < .001$). The effect sizes for specific language skills were as follows: Vocabulary has the highest effect size ($g = 1.32$), followed by Language in general ($g = .77$), Speaking ($g = .68$), Writing ($g = .63$), Reading ($g = .43$), and Listening ($g = .25$). These findings suggest that AI chatbots have the most substantial impact on learners' vocabulary acquisition, while they may enhance speaking and writing skills more effectively than reading and listening.

AI chatbots also produced a strong positive impact on learners' psychological and emotional aspects related to EFL learning ($g = 1.19 > 0.8$), with significant differences among groups ($p < .01$). The specific effect sizes are as follows: Learning Motivation ($g = 2.10$), Perceived Value ($g = 1.56$), Self-Efficacy ($g = 1.35$), Foreign Language Enjoyment ($g = 1.10$), and Learning Interest ($g = .96$). These results indicate that AI chatbots significantly improve learners' motivation, perceived value, self-efficacy, foreign language enjoyment, and interest in EFL study (Table 4).

4.2. Regulatory effect analysis

To better understand the influence of potential moderating variables on learners' learning effects, this study examined factors such as Intervention Durations, Educational Levels, Types of Teacher Roles, Types of Chabot Roles, and Types of Chabot Forms. The specific results are presented in Table 5.

The analysis revealed significant differences in the impact of varying intervention durations on learners' effects ($Q = 106.96$, $p < .001$). Long-term interventions demonstrated the highest effect size ($g = 1.21$), followed by medium-duration interventions ($g = 1.00$), while short-term interventions produced the lowest effect size ($g = .95$). These findings suggest a linear relationship between the length of the intervention and learning effects: the longer the intervention, the more substantial the improvement in the learner's results.

The results indicated that AI chatbots positively impacted learners across different educational levels. However, the differences between groups were not statistically significant ($Q = 3.62$, $p = .16 > .05$). The effect sizes, ranked from highest to lowest, were as follows: secondary education ($g = 1.94$), primary education ($g = 1.00$), higher education ($g = .77$).

The analysis shows that different teacher roles have notable effects on learners' effects, though the differences between groups are not statistically significant ($p = .85 > .05$). The largest effect size occurs when teachers primarily facilitate EFL learning ($g = 1.08$), while the effects value for teacher-assisted AI chatbots instruction and no teacher involvement are same ($g = .83$).

The analysis reveals significant differences in the impact of various robot roles on learners' effects ($Q = 94.38$, $p < .001$). The effect sizes for various robot roles are as follows: AI chatbots as a learning companion ($g = 1.13$), AI chatbots as a teaching assistant ($g = 1.03$), and AI chatbots as the primary instructor ($g = .77$). Therefore, AI chatbots are most effective when serving as learning companions or teaching assistants, while their effectiveness as primary instructors is lower.

Table 4. Testing AI chatbot's Impact on EFL.

Outcome Variables	N	n	Effect Size and 95% Confidence Interval			Heterogeneity Tests				Tau-Squared	Intergroup Effect		
			g	SE	95% CI	Q	df	P	I^2	τ^2	Q_B	df	P
Language Skill	2730	33	0.72	0.14	[0.44,1.01]	283.81	32	<.001	88.73	0.54	–	–	–
Listening	93	2	0.25	0.21	[-0.16,0.65]	0.82	1	0.365	0.00	0.00	21.25	5	0.001
Speaking	336	6	0.68	0.37	[-0.04,1.40]	49.28	5	<.001	89.85	0.72			
Reading	141	2	0.43	0.28	[-0.12,0.98]	1.71	1	0.191	41.47	0.08			
Writing	589	8	0.63	0.46	[-0.27,1.53]	104.38	7	<.001	93.29	1.36			
Vocabulary	160	5	1.32	0.47	[0.41,2.24]	26.97	4	<.001	85.17	0.91			
Language in General	1411	10	0.77	0.22	[0.34,1.20]	79.41	9	<.001	88.67	0.37			
Psychological/ Emotional	5139	37	1.19	0.15	[0.90,1.48]	640.53	36	<.001	94.38	0.67	–	–	–
Learning Motivation	700	7	2.10	0.56	[1.00,3.20]	165.36	6	<.001	96.37	1.72	16.49	4	0.002
Learning Interest	1630	12	0.96	0.25	[0.52,1.48]	207.25	11	<.001	94.69	0.65			
Perceived Value	784	5	1.56	0.53	[0.51,2.60]	70.45	4	<.001	94.32	1.07			
Learning Self-Efficacy	467	6	1.35	0.47	[0.42,2.28]	70.98	5	<.001	92.96	1.15			
Foreign Language Enjoyment	1558	7	1.10	0.30	[0.51,1.69]	110.00	6	<.001	94.55	0.55			

Notes: N, total number of participants; n, number of effect sizes; g, Hedges' g; SE, standard error; CI, confidence interval; Q, Cochran's Q statistic; df, degrees of freedom; P, p value; I^2, heterogeneity percentage; τ^2, between-study variance.
$*p < .05; **p < .01; ***p < .001$.

Table 5. Summary of regulatory effect analysis.

Adjustment variables	Category of variables	N	n	Effect size and 95% confidence interval			Heterogeneity tests				Tau-squared	Intergroup effect		
				g	SE	95% CI	Q	df	P	I²	τ²	Q_B	df	P
Intervention durations	Short	738	21	0.95	0.13	[0.70,1.21]	142.57	20	0.00	85.97	0.261	106.96	3	<.001
	Intermediate	1041	15	1.00	0.24	[0.53,1.46]	246.30	14	0.00	94.32	0.79			
	Long	2094	31	1.21	0.20	[0.81,1.61]	467.13	30	0.00	93.58	0.10			
	Not specified	306	3	–	0.09	[-0.18,0.19]	0.00	2	1.00	0.00	0.00			
Educational levels	Primary	1501	23	1.00	0.13	[0.75,1.26]	214.27	22	0.00	89.73	0.30	3.62	2	0.16
	Secondary	735	15	1.94	0.43	[1.10,2.77]	568.97	14	0.00	97.54	2.2			
	Tertiary	1943	32	0.77	0.12	[0.53,1.00]	176.11	31	0.00	82.40	0.33			
Types of teacher role	Tutor	1761	36	1.08	0.15	[0.79,1.37]	531.31	35	0.00	93.41	0.64	0.33	2	0.85
	Assistant	858	12	0.83	0.33	[0.19,1.46]	307.66	11	0.00	96.43	1.19			
	No teachers	1560	22	0.83	0.15	[0.54,1.12]	123.67	21	0.00	83.02	0.34			
Types of robot role	Teacher	562	22	0.77	0.19	[0.40,1.14]	134.13	21	0.00	84.34	0.59	94.38	2	<.001
	Assistant	2909	38	1.03	0.14	[0.75,1.31]	506.64	37	0.00	92.70	0.63			
	Peer	708	10	1.13	0.25	[0.65,1.61]	227.82	9	0.00	96.05	0.57			
Types of robot form	Anthropomorphic	1434	10	0.74	0.16	[0.42,1.06]	135.67	9	0.00	93.37	0.24	11.79	2	.003
	Zoomorphic	408	4	0.40	0.17	[0.07,0.73]	8.02	3	0.05	62.57	0.07			
	Mechanomorphic	2337	56	1.17	0.15	[0.87,1.47]	815.22	55	0.00	93.25	1.07			

Notes: categories with data sample size less than three are not included in the meta-analysis calculation; N, total number of participants; n, number of effect sizes; g, Hedges' g; SE, standard error; CI, confidence interval; Q, Cochran's Q statistic; df, degrees of freedom; P, p value; I², heterogeneity percentage; τ², between-study variance.
*$p < .05$; **$p < .01$; ***$p < .001$.

The analysis shows significant differences in the impact of robot appearance on learners' effects ($Q = 11.79$, $p > .01$). Specifically, AI chatbots with mechanical appearances have the largest effect size ($g = 1.17$), followed by human-like appearances ($g = .74$), and animal or cartoon appearances, which have the smallest effect size ($g = .40$).

5. Discussion

5.1. Impact of AI chatbots on EFL learning

This study finds that AI chatbots positively impact learners' EFL learning effects, consistent with the previous meta-analyses conducted by Wu and Yu (2024) and Wu and Li (2024). The positive effects are observed both in language skills and in psychological and emotional aspects of learning.

In terms of language skills, AI chatbots have shown positive effects on learners' listening ($g = .25$), speaking ($g = .68$), reading ($g = .43$), writing ($g = .63$), vocabulary ($g = 1.32$), and overall language abilities ($g = .77$), with significant differences between groups. Similarly, Wu and Li (2024) found that robot-assisted language learning enhances learners' listening, speaking, vocabulary, and overall language proficiency, although the effect on reading was less pronounced. This discrepancy may be attributed to chatbots' ability to provide real-time interaction and personalized feedback (Kwon et al., 2023; Song & Song, 2023), which not only increase learner engagement but also reinforce learners' correct language usage through immediate error correction. This process enables learners to consolidate their knowledge more effectively in real-world applications. Additionally, chatbots simulate authentic conversational scenarios, which improves learners' fluency and accuracy in language production (Hwang et al., 2022). However, research shows that improvement in receptive language skills (e.g. listening and reading) is less significant than productive skills (e.g. speaking and writing). This may be attributed to the complexity of language input, as listening and reading require extensive input and rich contexts (White, 1987), which the conversational capabilities of chatbots may not fully support. These findings suggest that future technological developments should focus more on integrating multimodal materials to enhance learning effects. Moreover, listening decoding skills are crucial for successful EFL comprehension, yet this area remains underexplored. Future research should pay more attention to this skill in AI chatbots-assisted EFL learning. Furthermore, with the rise of generative AI like ChatGPT, chatbots are increasingly being used for spoken interactions and writing feedback. However, fewer studies ($n = 2$) have focused on reading and listening, which may lead to bias in the statistical results.

In terms of psychological and emotional factors, the study finds that AI chatbots significantly enhance learners' motivation, self-efficacy, perceived value, and enjoyment of foreign language learning, while also stimulating and sustaining learners' interest in the subject. These findings are consistent with the study by Wu and Yu (2024).

First, regarding learning motivation, the study shows that AI chatbots effectively boost students' motivation, consistent with findings from Fryer and Carpenter (2006), Silitonga et al. (2023), and Wu and Yu (2024). The timely interaction and personalized feedback provided by AI chatbots create a more natural and authentic language learning environment, significantly improving EFL learners' motivation compared to traditional learning models. According to self-determination theory, autonomous dialogue can enhance intrinsic motivation, while reward-based prompts stimulate extrinsic motivation. In EFL learning, AI chatbots can trigger intrinsic motivation through autonomous conversations and boost extrinsic motivation through external rewards.

Second, regarding learning interest, this study aligns with the findings of Wu and Yu (2024) and Fryer et al. (2019), showing that AI chatbots enhance learners' interest in language learning. Fryer et al. (2019) compared the effects of chatbots and human learning partners on university students' interest in language learning (Jin & Zhang, 2021). The results indicated that AI chatbots, as learning partners, foster curiosity and engagement through interactive experiences. Additionally, the interactive and gamified features of AI chatbots further enhance learners' participation in the learning process.

Third, in terms of perceived value, this study finds that AI chatbots significantly increase students' perceived value, corroborating the findings of Hong et al. (2023). Integrating chatbots into gamified learning environments can enhance learners' perceived value and enjoyment of foreign language learning, thereby promoting vocabulary acquisition (Hong et al., 2023). The learning scenarios and social interactions facilitated by AI chatbots foster a stronger sense of social presence and perceived value, helping students to form deeper connections with the learning content.

Fourth, regarding self-efficacy, the results are consistent with Wu and Yu (2024) and Huang et al. (2022), indicating that AI chatbots significantly boost learners' self-efficacy. When students feel capable of completing a task, they are more likely to remain engaged (Bandura, 1977). Studies have shown that ChatGPT enhances learners' confidence and improves their speaking skills (Muniandy & Selvanathan, 2024). Compared to traditional learning methods, AI chatbots provide more comprehensive learning support and feedback, strengthening learners' self-efficacy.

Fifth, regarding foreign language enjoyment, this study confirms the findings of Jin and Zhang (2021) and Lee et al. (2023), showing that AI

chatbots significantly impact learners' enjoyment of foreign language learning (Kwon et al., 2023; Yang et al., 2022). The learning environments created by AI chatbots help alleviate the anxiety and nervousness associated with real-life interactions, enabling learners to engage in foreign language learning in a relaxed and comfortable manner.

Overall, AI chatbots positively impact the emotional and psychological aspects of EFL learning, though these effects are multifaceted and may be influenced by various factors (Jeon, 2024). As AI chatbots are still relatively new in educational settings, research in this area remains exploratory. Future research should continue to investigate the potential of AI chatbots to enhance foreign language learning, including incorporating educational games to increase learner participation and interest, and exploring the effects of group collaboration on learners' emotional experiences in AI chatbots-assisted foreign language learning.

5.2. Moderating factors in AI chatbots-Assisted EFL

The second key question of this study explores how potential moderating factors influence the effectiveness of AI chatbots in assisting EFL learning. The results indicate that the duration of intervention, chatbot role, and chatbot appearance are essential moderating factors.

The duration of the intervention significantly impacts learners' EFL effects, with longer intervention periods leading to better learning effects, consistent with findings by Alemdag (2023) and Wu & Li, (2024). This result is unsurprising, as learners are generally less familiar with AI chatbots compared to traditional teaching methods. However, as the duration increases, learners have more time to adapt to the technology, improve their proficiency, and subsequently improve their performance. Conversely, a recent meta-analysis found that short-term interventions had a more pronounced impact on learning effects (Wu & Yu, 2024), suggesting that novelty effects may play a significant role in the short term but diminish over time. Therefore, it is important to further investigate the role of novelty effects and explore the mechanisms by which different intervention durations influence learning effects.

The study also finds that AI chatbots significantly benefit learners across different educational levels, indicating that they can effectively support EFL learning at various stages (Alemdag, 2023; Wu & Li, 2024; Wu & Yu, 2024). Additionally, this study confirms that AI chatbots have a greater impact on K-12 learners than those in higher education (Alemdag, 2023; Wu & Li, 2024). This discrepancy may arise because higher education learners tend to possess more mature cognitive abilities and focus more on the practical applications of technology, while earlier generations of chatbots may not have met their needs as effectively. In

contrast, K-12 learners engage with simpler learning tasks and are more easily attracted by the entertaining features of AI chatbots (Wu & Li, 2024). Moreover, this phenomenon may reflect 'Marginal Effects', where additional technological inputs result in progressively smaller gains (Liu et al., 2024). In this study, younger learners, who had a weaker foundation in EFL, benefited more from the introduction of technology, while university students, starting with higher proficiency, experienced relatively smaller gains. Future research could adopt both cross-sectional and longitudinal comparisons to further investigate the impact of AI chatbots across different educational stages.

Regarding the role of teachers, this study found that learner effects were highest when teachers assumed a leading role. However, the differences in learning effects across various teacher roles did not reach statistical significance, consistent with the findings of Wu & Li, (2024). While AI chatbots support learners in studying foreign languages anytime and anywhere (Chen et al., 2020; Fryer & Carpenter, 2006; Kwon et al., 2023; Rad et al., 2023), it also require higher levels of information literacy and self-regulated learning (SRL) skills from learners. Learners with lower information literacy may struggle to use AI tools effectively without teacher assistance (Wu & Li, 2024). Although technology provides learners with greater autonomy, this study suggests that reliance solely on technology for autonomous learning is less effective. Teacher guidance remains essential, especially for developing higher-order skills and solving complex tasks. Interestingly, this study also found no significant difference between teacher-assisted learning and independent learning with AI chatbots.

The study reveals that the role of AI chatbots significantly affects learners' performance. Specifically, learners perform best when the AI chatbots function as a learning companion. This could be attributed to the higher levels of foreign language anxiety experienced when interacting with human peers (Guo et al., 2022; Hwang et al., 2022; Jeon, 2024; Wang et al., 2022; Zhang & Han, 2021), which can trigger negative emotions such as tension and unease (Wang et al., 2022). In contrast, AI chatbots provide personalized feedback and human-like interactions (Kwon et al., 2023; Tai & Chen, 2023; Wang et al., 2022), effectively reducing peer pressure and alleviating learners' foreign language anxiety, thus enhancing their performance. However, Wu & Li, (2024) found that AI chatbots as learning companions did not significantly impact learners' outcomes, suggesting that the effectiveness of AI chatbots may vary depending on the learning environment and the learners' backgrounds. Additionally, the study found that learners performed worst when AI chatbots assumed a tutor role (Wu & Li, 2024). This may stem from the current limitations of AI technology; despite rapid advancements, AI chatbots are still relatively

'weak' tools for language learning and possess considerable potential for improvement (Fryer et al., 2019). The previous analysis of teacher roles also indicated that teacher-led instruction yields the best effects, further supporting the notion that AI chatbots are more suitable as learning companions rather than independent tutors in collaborative teaching environments.

Regarding the appearance of AI chatbots, this study found that mechanical and humanoid AI chatbots had a profound impact on students' learning effects compared to those resembling animals or cartoon characters. Similar findings were reported by Wu & Li, (2024) and Wang et al. (2022), who demonstrated that the appearance of AI directly influences the frequency of use and the enjoyment learners derive from the experience. One possible explanation is that humanoid interactions from AI chatbots may better meet learners' social interaction needs, providing opportunities for communication and emotional expression (Hong et al., 2023). Additionally, the study found that context is critical for language learning (Chen et al., 2020; Guo et al., 2022). A humanoid appearance creates a more authentic interactive environment, allowing learners to engage in language exchanges without peer pressure, thus boosting confidence, interest, and motivation. In contrast, AI chatbots resembling animals or cartoon characters may lack the realism to create an authentic communication context. Interestingly, the study also revealed that a more realistic humanoid appearance does not necessarily lead to better learning effects. In some cases, overly realistic appearances may trigger 'The Uncanny Valley' effect, where the AI or robot's appearance is so close to human that it causes feelings of unease and aversion in learners (Dobson et al., 2018). Therefore, future research could explore the effects of different chatbot appearances and levels of anthropomorphism on learning effects, providing valuable insights for technology designers.

6. Implications and future directions

Based on the findings of this study and the analysis of existing literature, future applications and research on AI chatbots -assisted EFL learning should focus on the following areas for educators, instructional implementers, and technology developers:

6.1. Implications for instructional implementers

Firstly, based on the findings of this study, AI chatbots significantly enhance both students' language learning skills, as well as their psychological and emotional well-being. Therefore, educators should actively embrace intelligent technologies, master AI tools, and integrate them strategically into EFL instruction to improve learners' language abilities

and promote positive psychological and emotional experiences related to language learning.

Secondly, the study indicates that AI chatbots are most effective when used as learning companions or assistants. Teachers should consider incorporating chatbots into the learning process, to support autonomous learning, inquiry-based activities, and collaboration on complex tasks. This approach can significantly enhance students' learning effects.

Finally, While AI chatbots exhibit advanced capabilities and can replace teachers in some areas, this study reveals that students achieve better learning effects when teachers take a leading role in AI chatbots-assisted instruction. The 2024 report from the U.S. Office of Educational Technology, *'Artificial Intelligence and the Future of Teaching and Learning'* emphasizes the importance of human oversight in AI-based education: "We can call upon all stakeholders to adopt 'human in the loop' as a key criterion for educational use of AI" (Cardona et al., 2023). Therefore, teachers should remain aware of their crucial role in guiding learners, even when utilizing AI chatbots. When designing AI chatbots-based instruction, teachers should focus on leveraging the strengths of AI chatbots in language learning and integrate appropriate supportive activities. They should take the lead in developing tasks, activities, and assessments that foster the development of higher-order skills, such as critical thinking, creative thinking, problem-solving abilities and collaboration. Additionally, they should provide a range of learning scaffolds to assist learners in overcoming challenges and completing tasks effectively.

6.2. Recommendations for educational researchers

Firstly, current research shows that AI chatbots are particularly effective in enhancing speaking and writing skills. Future studies should focus on exploring their impact on other language skills, especially listening and reading (Li, 2022), to gain a comprehensive understanding of AI chatbots' effectiveness across all language domains and maximize their potential in language learning.

Secondly, related to the impact of intervention duration on learning effects remain inconsistent. Future research should investigate the specific mechanisms through which different intervention durations affect learning effects, including the technology adaptation processes and the duration and decay of novelty effects (Fryer et al., 2019). Understanding these dynamics can help optimize intervention strategies, ensuring AI chatbots remain effective at various stages of learning.

Thirdly, the role of teachers in influencing student learning warrants further examination, particularly in relation to factors such as sample size, educational level, and learning environment. Fourthly, future research

should consider additional variables that may affect learning effects, including educational stage (Alemdag, 2023; Wu & Li, 2024; Wu & Yu, 2024), gender (Jeon, 2024), learning methods (Chen et al., 2020), learners' initial skill levels, learning needs, and the complexity of teaching content. A comprehensive analysis of these factors will help develop personalized AI chatbots that cater to the specific needs and preferences of diverse learners.

6.3. Recommendations for technology developers

Firstly, technology developers should focus on integrating emotion recognition capabilities into AI chatbots for EFL. By using data to detect students' emotional states in real-time and adjusting feedback and interaction methods accordingly (Chang et al., 2010), developers can examine the impact of these adaptive features on students' learning effects and emotional experiences. Secondly, studies show that AI chatbots with mechanical and human-like appearances result in better learning effects. Designers should consider how the appearance of AI chatbots affects learners' emotions and interaction experiences (Wang et al., 2022; Wu & Li, 2024). Developing visually appealing and user-friendly designs can increase learner engagement and motivation, ultimately improving learning effects. Lastly, technology developers need to continually refine and update educational AI products to improve its understanding of students' input and learning needs, providing efficient feedback and timely responses (Wu & Li, 2024).

7. Conclusion

7.1. Major findings

The research systematically reviews empirical studies on the application of AI chatbots in the EFL, using meta-analysis to assess peer-reviewed articles published from 2022 to 2024. The study finds that AI chatbots significantly enhance learners' EFL learning effects, improving both their language skills, and psychological and emotional well-being. The analysis of moderating factors reveals that intervention duration, chatbot role, and chatbot appearance are critical variables. However, learners' educational level and teacher role did not show significant effect. Specifically, longer intervention periods resulted in better learning effects; chatbots acting as learning companions proved more effective; K12 learners showed greater improvement than higher education students; teacher-led instruction yielded better results compared to teacher-assisted or independent learning; and chatbots with mechanical appearances were more effective in improving learners' performance.

7.2. Major contributions

This study makes several important contributions. First, systematically analyzing empirical research from 2022 to 2024, provides up-to-date insight into the use of AI chatbots for EFL learning, offering timely and relevant findings. Second, the study adopts a holistic approach, examining the impact of AI chatbots not only on learners' language skills but also on their psychological and emotional states. Lastly, the research outlines future directions for further study in this field and provides practical recommendations for educators and technology developers, contributing to both theoretical advancement and educational practice.

7.3. Limitations

However, this study has limitations. First, while following PRISMA guidelines, it was challenging to retrieve and include all relevant literature, limiting the scope to 25 studies, which may affect the accuracy and reliability of the results. Future research should expand data sources and employ different research methods to more accurately assess the effectiveness of AI chatbots in EFL. Second, some studies lacking statistical data could not be included in the meta-analysis. This research only analyzed the impact of five moderator variables: intervention duration, learner academic level, teacher role, chatbot role, and chatbot type. Future studies could explore other potential variables, such as gender, cultural background, proficiency level, and learning methods.

Disclosure statement

Funding

Notes on contributors

Zhaoyang Liu is a graduate student in the Faculty of Education at Shaanxi Normal University. Her research interests focus on the application of intelligent technology in education.

Wenlan Zhang is a professor in the Faculty of Education at Shaanxi Normal University. Her research interests include the educational applications of intelligent technology, learning psychology and technology, and project-based learning.

Panpan Yang is a doctoral candidate in the Faculty of Education at Shaanxi Normal University. Her research interests include English teacher education and the application of intelligent technology in education.

ORCID

Zhaoyang Liu 🔟 http://orcid.org/0009-0008-6142-1710
Wenlan Zhang 🔟 http://orcid.org/0000-0001-9854-2307
Panpan Yang 🔟 http://orcid.org/0009-0002-1768-697X

References

Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning, 40*(5), 2398–2398. https://doi.org/10.1111/jcal.12948

Alemdag, E. (2023). The effect of chatbots on learning: A meta-analysis of empirical research. *Journal of Research on Technology in Education*, 1–23. https://doi.org/10.1080/15391523.2023.2255698

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research & Therapy*, *1*(4), 139–161. https://doi.org/10.1016/0146-6402(78)90002-4

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. *Office of Educational Technology*. https://coilink.org/20.500.12592/rh21zz

Chang, C. W., Lee, J. H., Wang, C. Y., & Chen, G. D. (2010). Improving the authentic learning experience by integrating robots into the mixed-reality environment. *Computers & Education*, *55*(4), 1572–1578. https://doi.org/10.1016/j.compedu.2010.06.023

Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A ChatBot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, *58*(6), 1161–1189. https://doi.org/10.1177/0735633120929622

Chen, H., Liu, X., Yin, D., & Tang, J. (2017). A survey on dialogue systems: Recent advances and new frontiers. *Acm Sigkdd Explorations Newsletter*, *19*(2), 25–35. https://doi.org/10.1145/3166054.3166058

Chien, Y. C., Wu, T. T., Lai, C.-H., & Huang, Y.-M. (2022). Investigation of the influence of artificial intelligence markup language-based LINE ChatBot in contextual English learning. *Frontiers in Psychology*, *13*, 785752. https://doi.org/10.3389/fpsyg.2022.785752

Cooper, H., Hedges, L. V., Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. https://muse.jhu.edu/pub/207/edited_volume/book/65827

Dobson, N., Roe, A. H., Ratelle, A., & Ruddell, C. (Eds.). (2018). *The animation studies reader*. Bloomsbury Academic. https://doi.org/10.5040/9781501332647

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language, Learning and Technology*, *10*, 8–14.

Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, *120*, 103–209. https://doi.org/10.1016/j.system.2023.103209

Görgen, R., Huemer, S., Schulte-Körne, G., & Moll, K. (2020). Evaluation of a digital game-based reading training for German children with reading disorder. *Computers & Education*, *150*, 103834. https://doi.org/10.1016/j.compedu.2020.103834

Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, *54*, 100666. https://doi.org/10.1016/j.asw.2022.100666

Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers & Education*, *179*, 104395. https://doi.org/10.1016/j.compedu.2021.104395

Hedges, L., Rosenbaum, P. R., & Olkin, I. (1987). *Statistical methods for meta-analysis* (vol. 82). Academic Press. https://www.jstor.org/stable/2289186?origin=crossref

Hong, J. C., Lin, C. H., & Juh, C. C. (2023). Using a Chatbot to learn English via Charades: The correlates between social presence, hedonic value, perceived value, and learning outcome. *Interactive Learning Environments*, *32*(10), 6590–6606. https://doi.org/10.1080/10494820.2023.2273485

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.12610

Hwang, W.-Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*, *37*(7), 1632–1658. https://doi.org/10.1080/09588221.2022.2095406

Jeon, J. (2023). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, *36*(7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272

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

Jin, Y., & Zhang, L. J. (2021). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *International Journal of Bilingual Education and Bilingualism*, *24*(7), 948–962. https://doi.org/10.1080/13670050.2018.1526253

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second language learning chatbot. *Language Learning*, *26*(1), 1–18.

Kim, N. Y. (2022). English with AI: A new era of TOEIC learning for students majoring in airline services. *Linguistic Research*, *39*(3), 97–122. https://doi.org/10.17250/KHISLI.39.202209.004

Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology*, *27*(1), 1–19.

Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, *54*(2), 508–516. https://doi.org/10.1177/00336882231165060

Lee, J. H., Yang, H., Shin, D., & Kim, H. (2020). Chatbots. *ELT Journal*, *74*(3), 338–344. https://doi.org/10.1093/elt/ccaa035

Lee, Y. J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies*, *14*(1), e202412. https://doi.org/10.30935/ojcmt/14195

Li, R. (2022). Effects of mobile-assisted language learning on EFL/ESL reading comprehension. *Educational Technology & Society*, *25*(3), 15–29.

Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16–35.

Liu, Z., Zhang, W., & Hu, L. (2024). Can desktop virtual reality effectively enhance academic achievement?—A meta-analysis. *Innovations in Education and Teaching International*, 1–16. https://doi.org/10.1080/14703297.2024.2354753

Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning*, *29*(3), 451–476. https://doi.org/10.1080/09588221.2014.963123

Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, *33*(8), 865–890. https://doi.org/10.1080/09588221.2019.1595664

Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a Public University, Malaysia. *Teaching Public Administration*, 01447394241230152. https://doi.org/10.1177/01447394241230152

Newton, J. M., Ferris, D., Goh, C., Grabe, W., Stoller, F., & Vandergrift, L. (2018). *Teaching English to second language learners in academic contexts*. Routledge. https://doi.org/10.4324/9781315626949

OpenAI. (2023). *GPT-4 system card.* https://cdn.openai.com/papers/gpt-4-system-card.pdf

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., … Mulrow, C. D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Polyzi, P., & Moussiades, L. (2023). An artificial vocabulary learning assistant. *Education and Information Technologies*, *28*(12), 16431–16455. https://doi.org/10.1007/s10639-023-11810-9

Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, *32*(9), 5020–5040. https://doi.org/10.1080/10494820.2023.2208170

Randall, N. (2020). A survey of robot-assisted language learning (RALL). *ACM Transactions on Human-Robot Interaction*, *9*(1), 1–36. https://doi.org/10.1145/3345506

Ryong, K., Lee, D., & Lee, J. (2023). Chatbot's complementary motivation support in developing study plan of E-learning English lecture. *International Journal of Human–Computer Interaction*, *40*(10), 2641–2655. https://doi.org/10.1080/10447318.2022.2163786

Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, *13*(9), 1937–1960. https://doi.org/10.3390/ejihpe13090140

Silitonga, L. M., Hawanti, S., Aziez, F., Furqon, M., Zain, D. S. M., Anjarani, S., & Wu, T. T. (2023). The impact of AI chatbot-based learning on students' motivation in English writing classroom. In *International conference on innovative technologies and learning* (pp. 542–549). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-40113-8_53

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, *37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28. https://doi.org/10.1145/357980.357991

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence1. *Applied Linguistics*, *8*(2), 95–110. https://doi.org/10.1093/applin/8.2.95

Wu, R., & Yu, Z. (2024). Do AI Chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Wu, X., & Li, R. (2024). Effects of robot-assisted language learning on English-as-a-foreign-language skill development. *Journal of Educational Computing Research*, 1–25. https://doi.org/10.1177/07356331231226171

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/S0958344022000039

Ye, Y., Deng, J., Liang, Q., & Liu, X. (2022). Using a smartphone-based Chatbot in EFL Learners' oral tasks. *International Journal of Mobile and Blended Learning (IJMBL)*, *14*(1), 1–17. https://doi.org/10.4018/IJMBL.299405

Yuan, Y. (2023). An empirical study of the efficacy of AI Chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments*, *32*(10), 6774–6789. https://doi.org/10.1080/10494820.2023.2282112

Zhang, H., & Han, X. (2021). Influence of vocalized reading practice on English learning and psychological problems of middle school students. *Frontiers in Psychology*, *12*, 709023. https://doi.org/10.3389/fpsyg.2021.709023

Zhang, R., Zou, Di., & Cheng, G. (2023). Chatbot-based training on logical fallacy in EFL argumentative writing. *Innovation in Language Learning and Teaching*, *17*(5), 932–945. 10.1080/17501229.2023.2197417

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, *15*(4), 2872. https://doi.org/10.3390/su15042872