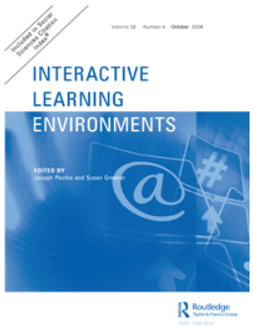

Interactive Learning Environments

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/nile20

A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models

Jaeho Jeon, Seongyong Lee & Seongyune Choi

To cite this article: Jaeho Jeon, Seongyong Lee & Seongyune Choi (2024) A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models, Interactive Learning Environments, 32:8, 4613-4631, DOI: 10.1080/10494820.2023.2204343

To link to this article: https://doi.org/10.1080/10494820.2023.2204343

INTERACTIVE LEARNING ENVIRONMENTS
2024, VOL. 32, NO. 8, 4613–4631
https://doi.org/10.1080/10494820.2023.2204343

Routledge
Taylor & Francis Group

REVIEW ARTICLE

A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models

Jaeho Jeon [a], Seongyong Lee [b] and Seongyune Choi [c]

[a]Department of Literacy, Culture, and Language Education, Indiana University, Bloomington, IN, USA; [b]Department of English Education, Hannam University, Daejeon, Republic of Korea; [c]Department of Computer Science and Engineering, Korea University, Seoul, Republic of Korea

ABSTRACT

Chatbot research has received growing attention due to the rapid diversification of chatbot technology, as demonstrated by the emergence of large language models (LLMs) and their integration with automatic speech recognition. However, among various chatbot types, speech-recognition chatbots have received limited attention in relevant research reviews, despite their increasing potential for language learning. To fill this gap, 32 empirical studies on speech-recognition chatbots for language learning were reviewed. The following information was reviewed for each study: basic publication information, research focus, location of chatbot use, methodology, group design format, participant information, intervention duration, target language, device type adopted, and chatbot role. An upward trend in research quantity starting in 2020 was identified, which accelerated exponentially in 2022. College students were more likely than other groups to be involved in research, and English as a second or foreign language was the most common target language. Most studies focused on participants' perceptions of chatbots and the degree to which using chatbots helped them develop their speaking or listening proficiency. Methodologically, single-chatbot design using mixed methods was the most common design format, and most studies were conducted for more than one month in laboratory or classroom settings. Conventional mobile devices, such as smartphones, tablet PCs, and smart speakers without a screen, were the most frequently adopted device types. The chatbots' most common role was as conversational partner. A detailed discussion of these results and their implications for future research on speech-recognition chatbots, particularly regarding the use of LLM-powered chatbots, is provided.

ARTICLE HISTORY
Received 17 January 2023
Accepted 11 April 2023

KEYWORDS
Chatbot; large language model; ChatGPT; artificial intelligence; automatic speech recognition; intelligent personal assistant; language learning; computer-assisted language learning

1. Introduction

A chatbot is a virtual agent that can communicate with users in natural language in text and/or audio mode. In line with the advances in artificial intelligence (AI) technologies, chatbot technology is also becoming more sophisticated and human-like. Historically, chatbots were initially developed as text-only agents on desktop computers (Weizenbaum, 1966). However, they have been enhanced to incorporate current technologies such as mobile devices, natural language processing (NLP) capabilities, and automatic speech recognition (ASR), making chatbots more accessible in everyday life and more human-like in their interactions (Lee & Jeon, 2022; Shadiev & Liu, 2023). While this

CONTACT Seongyong Lee seongyong@hnu.kr, seongyonglee77@gmail.com

evolutionary trajectory has advanced consistently, however, its momentum has recently accelerated remarkably owing to the rapid diversification of chatbot technology. Among its notable advances are the emergence of large language models (LLMs), such as OpenAI's GPT series, Google's Bard, and Meta's BlenderBot, and their integration with complementary technologies, including ASR and computer vision (OpenAI, 2023; Rospigliosi, 2023; Zhang, 2023). The recent development of applications such as "Talk-to-ChatGPT" and Google's Universal Speech Model (Zhang, 2023) represents a significant step forward in spoken human-chatbot communications. Accordingly, people increasingly communicate with chatbots for a variety of purposes, including the learning and teaching of languages (Fryer et al., 2020; Jeon, 2022b).

In light of this changing communication landscape, language researchers have paid renewed attention to chatbots, particularly those with speech-recognition functionality. For example, they have examined how language learners respond to communication breakdowns with chatbots (Moussalli & Cardoso, 2020); how interactions with chatbots influence their learning outcomes (Dizon, 2020; Tai & Chen, 2022a); and how they perceive the effects of using chatbots on such affective variables as willingness to communicate, motivation, and anxiety (Chen et al., 2020; Jeon, 2022b; Tai & Chen, 2020). Overall, these studies have confirmed the benefits of using speech-recognition chatbots for language learning. Also, as various LLM-powered chatbots with speech-recognition functionality become increasingly prevalent in society, they may be expected to attract increasing scholarly attention.

On the other hand, previous research on educational uses of chatbots indicates that their effectiveness for promoting student learning is not a straightforward issue. That is, several factors have emerged across studies that warrant attention from a global standpoint. These include learner-related factors (e.g., participants' educational levels, language proficiency, willingness to communicate, technology readiness), technology-related factors (e.g., device type, novelty effect, chatbot modality), and context-related factors (e.g., formal or informal location for chatbot use).

The literature does include some recent systematic and comprehensive analyses of chatbot studies. For instance, Huang et al. (2022), analyzed 25 empirical studies on the pedagogical, technological, and social affordances of both speech- and text-based chatbots utilized in language learning. For each category, they described the roles of chatbots, their language learning benefits, and summaries of sample studies. Drawing on 343 publications on conversational technology for language learning, Bibauw et al. (2019) introduced the term "dialogue-based Computer Assisted Language Learning (CALL)" to refer to "any system or application where the activity consists for the learner to engage in a dialogue with an automated interlocutor" (p. 863). They proposed a typology of four dialogue systems based on the extent to which a combination of form and meaning constraints was applied. Ji et al. (2023) conducted a systematic review of 24 empirical studies of both speech- and text-based chatbot usage in language education, focusing on collaboration between chatbots and human teachers. They found that in many studies, chatbots were used without teachers' involvement and concluded that promoting such collaboration might help to overcome numerous constraints inherent to chatbot technology.

Although these focused reviews have enriched the literature on chatbots for language learning, there remain two important gaps. First, in previous reviews, studies of speech- and text-based chatbots were addressed as one group. Given the rapid development of ASR and NLP technologies and the increasing attention to speech technology in language studies (Shadiev & Liu, 2023), different implications may be drawn from reviews of studies focused on speech-recognition chatbots and those in which chatbot modality was not distinguished. Second, previous reviews focused mainly on certain features of chatbot technology itself, such as affordances, chatbot typology, and chatbot-human relationships in language learning. This selective focus suggests that other aspects of chatbots are underexplored, particularly in comparison with the more comprehensive range of systematic reviews of chatbot research in adjacent disciplines, calling for a broader approach to the relevant literature in language education (e.g. Hwang & Chang, 2021; Liang et al., 2021). Lastly, despite growing societal interest in the educational potential of LLMs, these reviews have not provided direct insights into this aspect. A comprehensive analysis of extant research on

chatbots will yield valuable insights that both researchers and practitioners can utilize to further explore the educational possibilities of emerging generations of LLM-powered chatbots.

To address these concerns, we conducted a comprehensive systematic review of research on chatbots in language learning, with a particular focus on those with speech-recognition functionality. This analysis may serve as an important reference for future explorations of educational chatbots, including various types of LLM-powered chatbots. To examine previous studies, we used three overarching aspects comprising 11 specific categories:

- Publication aspects: basic publication information (publication year and journal).
- Research design aspects: research focus, research methodology, group design format, location of chatbot use, number of participants, participants' educational levels, target language, and duration of intervention.
- Technology aspects: device type adopted and role of chatbot.

2. Methods

2.1. Data collection

To conduct a systematic review, we used the analytical protocol of the PRISMA guidelines (Moher et al., 2009). Following Hwang and Chang (2021), we selected the Web of Science database as our data source, which also ensured the quality of the publications (Hwang & Tsai, 2011).

To decide search terms, we first conducted a preliminary manual search of the highly cited academic journals in the fields of educational technology and language learning. To identify a search string that sufficiently reflected the current trends in chatbot research, we reviewed recently published studies and systematic meta-analyses of research on conversational technology for language learning (e.g. Bibauw et al., 2022; Huang et al., 2022). Consequently, we used the search string shown in Table 1.

The search was completed by the beginning of December 2022. It was based on the title, abstract, and keywords (TITLE-ABS-KEY) and then filtered for peer-reviewed articles written in English. Based on the search results, we included an article if it (1) reported an empirical study on the use of chatbots in language learning; (2) focused on speech-recognition chatbots; (3) contained sufficient information about chatbot design; and (4) contained sufficient details about students' learning, experiences with, or perspectives on, chatbots.

Figure 1 shows the procedures for the literature search and selection process. Because we aimed to investigate the use of speech-recognition chatbots, we needed to screen out studies of chatbots that lacked this functionality. However, during the screening stage, we could not identify all of these studies simply by reviewing titles and abstracts. Therefore, at this stage, we excluded only studies for which the use of text-only chatbots was explicitly mentioned in the titles or abstracts.

During the next stage of determining eligibility, we reviewed the full text of each study to identify studies involving speech-recognition chatbots. We excluded articles that (1) explicitly indicated exclusive use of text-based chatbots with no speech-recognition functionality (e.g. Hew et al., 2023; Jeon,

Table 1. Keyword search scheme.

Technology		Language
"Chatbot*"	*AND*	"Second language*"
"Conversational agent*"		"Foreign language*"
"Pedagogical agent*"		"EFL"
"Chatterbot*"		"ESL"
"Conversational system*"		"Language learning"
"Dialogue system*"		"Language teaching"
"Spoken dialogue system*"		"Language acquisition"
"Intelligent personal assistant*"		

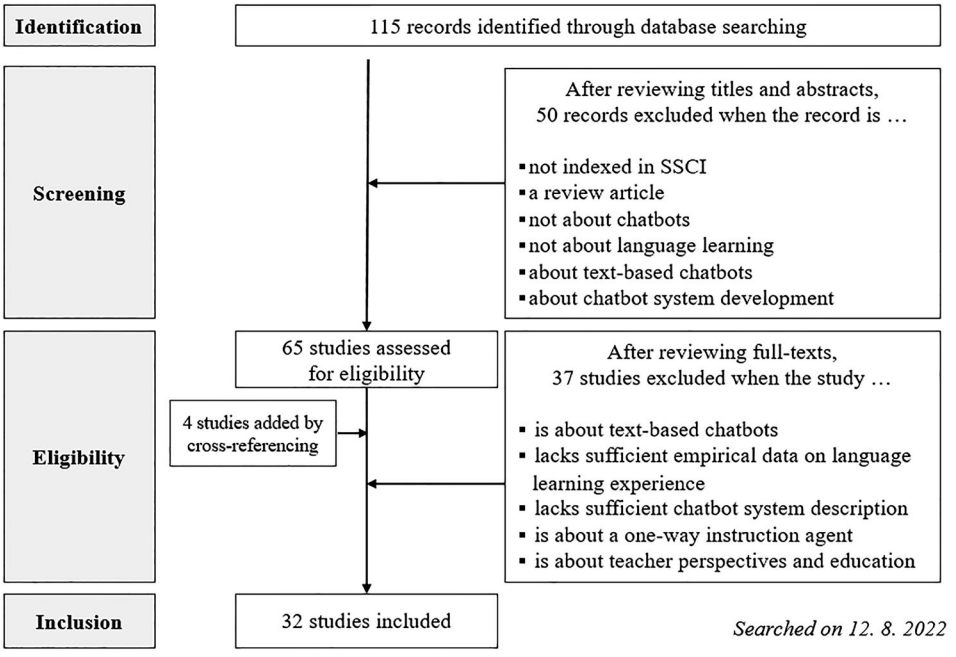

Figure 1. Literature search and selection process.

2021); (2) did not provide a clear description of chatbot operation modality or indicated the combined use of speech- and text-based chatbots (e.g. Hsu, 2022); (3) dealt with a one-way instruction agent, typically called a pedagogical agent (e.g. Davis et al., 2019); (4) did not present evidence of learners using speech-recognition functionality (e.g. Coniam, 2014); (5) represented teaching or professional development perspectives (e.g. Timpe-Laughlin, Sydorenko, Daurio, 2022); or (6) were not about learning but about the feasibility of using a chatbot system (e.g. van Doremalen et al., 2016). Finally, the references of the final set of studies and recent review studies (Bibauw et al., 2022; Huang et al., 2022; Ji et al., 2023) were checked to identify other potentially relevant studies, which resulted in the addition of four studies. In total, 32 studies were selected for this review.

2.2. Data analysis

For thematic data analysis, a hybrid approach that employed both deductive and inductive processes was taken to prepare a coding scheme (Braun & Clarke, 2006; Xu & Zammit, 2020). First, using each article as an analytical unit, we thoroughly and iteratively read all the articles and extracted all the information relevant to the publication, research design, and technology aspects of each study.

Next, we consulted previous systematic reviews focused on analyzing research trends and issues related to topics adjacent to speech-recognition chatbots. For example, to extract frameworks that could guide our deductive coding precedures, we examined studies on chatbots in general and in language education (Hwang & Chang, 2021; Ji et al., 2023), on AILEd (Artificial Intelligence in Language Education) (Liang et al., 2021), on ASR technology (Shadiev & Liu, 2023), and on mobile technology (Elaish et al., 2022; Fu & Hwang, 2018; Shadiev et al., 2017; Xie et al., 2019). We then consulted these frameworks to formulate the initial basis for coding. The result was a coding scheme of nine categories: basic publication information, research focus, methodology, number of participants, participants' educational level, duration of intervention, target language, device adopted, and chatbots' role in the learning process.

Table 2. Coding schemes.

	Category	Sub-category	Reference
Publication aspect	Basic information	• Year of publication • Journal names	Elaish et al. (2022)
Research design aspect	Research focus	• Perceptions - Perceptions of chatbot use - Affective variables - Cultural constructs • Proficiency - Listening - Speaking - Writing - Reading - Vocabulary - Idioms/phrases - Pronunciation - Grammar • Learning behaviors - Learning records on the system - Observation • Learning differences - Culture/country - Proficiency - Gender - Chatbot usage amount - Perception of chatbot system - Willingness to communicate	Shadiev et al. (2017), Liang et al. (2021)
	Location of chatbot use	• In class • Outside class • Both inside and outside classroom • In lab	N/A
	Methodology	• Quantitative • Qualitative • Mixed-methods	Hwang and Chang (2021)
	Group design format	• Single-chatbot design • Chatbot vs. human design • Chatbot vs. chatbot design • Chatbot vs. chatbot vs. human design • Chatbot vs. traditional material design	N/A
	Number of participants	• Fewer than 30 • Between 30 and 60 • More than 60	Shadiev and Liu (2023)
	Participants' educational level	• Higher education • Secondary school • Elementary school • Preschool • Working adults	Fu and Hwang (2018)
	Intervention duration	• Less than one week	Shadiev and Liu (2023)

(*Continued*)

Table 2. Continued.

	Category	Sub-category	Reference
		• Between one week and one month • More than one month	
	Target language	• L2 English • L1 English • L2 Chinese • L2 French	Shadiev and Liu (2023)
Technology aspect	Adopted device type	• Smartphone • Desktop PC • Tablet PC • Wearable device • Smart speaker with a screen • Smart speaker without a screen • Others	Xie et al. (2019)
	Chatbot role	• Conversational partner • Feedback provider • Evaluator • Resource provider • Interviewer	Ji et al. (2023)

However, we found that the combination of the referenced frameworks did not always sufficiently cover the extracted data or sometimes included irrelevant sub-categories. Therefore, we also performed a data-driven inductive coding process, by which we were able to identify two primary emergent categories of chatbot research: group design format and location of chatbot use. Further, we identified additional codable sub-categories in each main category, which we used to refine the existing frameworks in addition to deleting irrelevant sub-categories. For example, to code the adopted device type, we first referred to Xie et al.'s (2019) systematic review, which identified four device types: "wearable devices," "smartphones," "tablet computers," and "traditional computers or devices." However, we found that in our data set, many researchers had utilized "smart speakers with a screen display," "smart speakers without a screen display," and in a few cases "other types." Therefore, we included those three emergent types in our framework, which resulted in a total of seven sub-categories. Applying the same process to all the categories under investigation, the two researchers finalized the preparation of the coding scheme, with a total of 11 sub-categories, which were then classified into three overarching categories: publication, research design, and technology. The categories, their sub-categories, and the referenced frameworks are presented in Table 2.

The data obtained from the chosen papers were then coded independently by two of the researchers, one of whom was involved in the preparation of the coding scheme and one who was not. Their inter-coder reliability, which was calculated following the guideline for preventing coder bias, was 0.91, indicating high inter-coder reliability (O'Connor & Joffe, 2020).

3. Results

3.1. Publication aspects

3.1.1. Publication years

Figure 2 displays the distribution of years of publication of articles on speech-recognition chatbots for language learning. After the first article (Morton & Jack, 2010) was published, no article was found

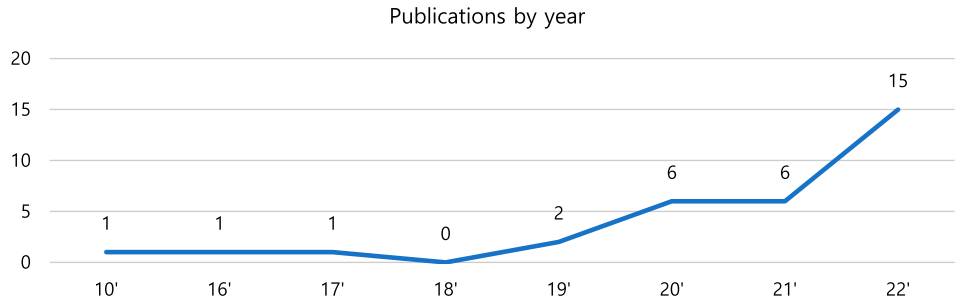

Figure 2. Numbers of publications by year.

Table 3. Publication quantity by journal.

Journal	*N*
Computer Assisted Language Learning	12
Interactive Learning Environment	8
Computers & Education	2
Computers in Human Behavior	2
Child Development	1
Intercultural Pragmatics	1
Frontiers in Psychology	1
Education and Information Technologies	1
Educational Technology & Society	1
ReCALL	1
IEEE Transactions on Learning Technologies	1
Language Learning & Technology	1
Total	32

from 2011 to 2015. From 2016 to 2019, the number of research publications remained low at around one or two articles per year with no article published in 2018. However, in 2020, with six publications, a rising trend in the number of published studies began, which then accelerated at an exponential rate to 15 publications in 2022.

3.1.2. Journals of publication

Table 3 presents the distribution of the articles by published journals. It was found that research on speech-recognition chatbots for language learning was published in a total of 12 different journals in the fields of educational technology (e.g. *Computers & Education*), linguistics (e.g. *Intercultural Pragmatics*), and adjacent fields such as psychology (e.g. *Child Development*). Among the journals, *Computer Assisted Language Learning* was found to have the most relevant articles ($n = 12$), which is not surprising as this journal addresses the two core topics of the current review: educational technology and language learning.

3.2. Research design aspects

3.2.1. Research focus

Table 4 presents the research focuses of the reviewed articles. In many studies, multiple aspects related to utilizing chatbots for language learning were examined. For example, Divekar et al. (2021) examined L2 Chinese learners' vocabulary, speaking, and listening proficiency test scores after they interacted with chatbots posing as shopkeepers at a Chinese street market. They also examined how the students perceived the chatbot and the degree to which their cultural understanding of Chinese street markets was enhanced.

Table 4. Research focus.

Research focus	Subcategory	Publication number
Perception	Perceptions of chatbot use	25
	Affective Variables	12
	Cultural constructs	1
	Total	38
Proficiency	Listening	9
	Speaking	11
	Writing	0
	Reading	2
	Vocabulary	3
	Pronunciation	2
	Grammar	1
	Total	28
Learner behaviors	Recorded logs	14
	Observation	2
	Total	16
Learner differences	Culture/country	2
	Proficiency	1
	Gender	1
	Chatbot usage amount	2
	Perception of chatbot system	1
	Willingness to communicate	1
	Total	8

According to our analysis, the most common research focus was on perceptions ($n = 35$), followed by language proficiency ($n = 28$), learner behaviors ($n = 16$), and learner differences ($n = 8$). Studies of perceptions focused on three aspects: learners' perceptions of chatbots in language learning contexts ($n = 25$); the effects of the use of chatbots on learners' affective variables, such as willingness to communicate, learning motivation, and confidence ($n = 12$); and cultural constructs ($n = 1$). Among the 28 studies that addressed language proficiencies, speaking ($n = 11$) and listening ($n = 9$) were the most common targets. In 14 of the 16 studies in which learner behavior was examined, automatically generated transcription logs of conversations between chatbots and learners were utilized to analyze different types of learner behavior or conversation patterns in learner-chatbot interactions. These included learners' communication strategies, chatbot-generated task success rates, and lengths of learners' responses (e.g. Chen et al., 2020; Yang, Kim, et al., 2022). In two studies, learner-chatbot interaction was observed by a researcher (e.g. Xu et al., 2021). Last, the eight studies that addressed learner differences focused on differences in how language learners with different characteristics experienced or benefitted from interactions with chatbots. Researchers used factors that have traditionally been used in previous literature, such as cultural/national differences ($n = 2$), proficiency levels ($n = 1$), willingness to communicate ($n = 1$), and gender ($n = 1$) as well as new factors relevant to chatbots, such as differences in the amount of chatbot usage ($n = 2$) and perceptions of chatbots ($n = 1$).

3.2.2. Research methodology

As shown in Table 5, in the majority of the studies reviewed, mixed-methodologies ($n = 21$) were used, followed by quantitative methods ($n = 10$). Only one study was qualitative (Jeon, 2022b). In most of the mixed-methods studies, a quasi-experimental pre-test/post-test design was used to measure and compare achievement between experimental and control groups. Students' perceptions of chatbots were examined using qualitative tools such as interviews or open-ended questionnaires (e.g. Tai, 2022; Tai & Chen, 2022b).

3.2.2. Group design format

Table 5 presents the different group design formats used. The most common was a single-chatbot group design ($n = 15$), in which researchers deployed only one chatbot type. The other formats were

Table 5. Elements of selected studies' research designs (*searched on 12. 08. 2022*).

Author	Year	Method	Group design format	Location of chatbot use	Participant number	Education level	Intervention duration	Target language
Ebadi and Amini	2022	Mixed	Single-chatbot group	Outside class	256	Higher Edu.	3 weeks	L2 English
Hwang et al.	2022	Mixed	Chatbot vs. chatbot	In class	43	Elementary	7 weeks	L2 English
Tai	2022	Mixed	Chatbot vs. human	Outside class	89	Higher Edu.	14weeks	L2 English
Lee and Jeon	2022	Mixed	Single-chatbot group	In lab	67	Elementary	Less than 1 week	L2 English
Chien et al.	2022	Quant	Chatbot vs. chatbot	Outside class	73	Secondary	4 weeks	L2 English
Tai and Chen	2022a	Mixed	Chatbot vs. chatbot vs. human	In class	92	Secondary	10 weeks	L2 English
Tai and Chen	2022b	Mixed	Chatbot vs. chatbot vs. human	In class	88	Secondary	10 weeks	L2 English
Yang, Kim, et al.	2022	Mixed	Single-chatbot group	In class	314	Elementary, secondary	1–3 weeks	L2 English
Yang, Lai, et al.	2022	Mixed	Chatbot vs. chatbot	Outside class	34	Higher Edu.	6 weeks	L2 English
Wang, Liu, et al.	2022	Mixed	Single-chatbot group	In- & outside class	16	Elementary	3 months	L2 English
Wang, Pang, et al.	2022	Quant	Single-chatbot group	In- & outside class	327	Elementary	2.5 month	L2 English
Jeon	2022a	Mixed	Single-chatbot group	Outside class	179	Elementary	8 weeks	L2 English
Jeon	2022b	Quali	Single-chatbot group	In class	36	Elementary	16 weeks	L2 English
Timpe-Laughlin, Sydorenko, and Dombi	2022	Mixed	Chatbot vs. human	In lab	47	Higher Edu.	Less than 1 week	L2 English
Xu et al.	2022	Quant	Chatbot vs. chatbot vs. human	In lab	117	Pre-school	Less than 1 week	L1 English
Hsu, Chen, and Todd	2021	Mixed	Chatbot vs. human	In class	50	Higher Edu.	7 weeks	L2 English
Lin and Mubarok	2021	Mixed	Chatbot vs. chatbot	In- & outside class	50	Higher Edu.	3 weeks	L2 English
Gonulal	2021	Mixed	Single-chatbot group	In lab	42	Higher Edu.	Less than 1 week	L2 English
Hsu, Chen, and Yu	2021	Quant	Chatbot vs. traditional material	Outside class	100	Higher Edu.	4 months	L2 English
Divekar et al.	2021	Mixed	Single-chatbot group	In lab	10	Higher Edu.	2 weeks	L2 Chinese
Xu et al.	2021	Quant	Chatbot vs. human	In lab	90	Pre-school	Less than 1 week	L1 English
Tai and Chen	2020	Mixed	Single-chatbot group	In class	112	Secondary	2 weeks	L2 English
Chen et al.	2020	Mixed	Single-chatbot group	In lab	29	Higher Edu.	Less than 1 week	L2 English
Ayedoun et al.	2020	Quant	Chatbot vs. chatbot	In lab	60	Higher Edu.	Less than 1 week	L2 English
Timpe-Laughlin and Dombi	2020	Mixed	Single-chatbot group	In lab	107	Higher Edu., working adults	Less than 1 week	L2 English
Dizon	2020	Quant	Chatbot vs. human	In class	28	Higher Edu.	10 weeks	L2 English
Moussalli and Cardoso	2020	Mixed	Single-chatbot group	In lab	11	Higher Edu.	Less than 1 week	L2 English
Forsyth et al.	2019	Mixed	Chatbot vs. human	In lab	31	Elementary	Less than 1 week	L2 English
Fryer et al.	2019	Mixed	Chatbot vs. human	In class	91	Higher Edu.	15 weeks	L2 English
Fryer et al.	2017	Quant	Chatbot vs. human	In class	122	Higher Edu.	12 weeks	L2 English
Hassani et al.	2016	Quant	Single-chatbot group	In lab	10	Higher Edu.	Less than 1 week	L2 English
Morton and Jack	2010	Quant	Single-chatbot group	In lab	76	Secondary	Less than 1 week	L2 French, L2 English

variations of comparison designs, in which different group conditions were created and compared. The most prevalent comparison design format involved comparisons between humans and chatbots as interlocutors ($n = 8$). For example, Forsyth et al. (2019) compared students' responses to being interviewed by a chatbot and by a human. Dizon (2020) compared the experimental group that used chatbots for part of in-class activities with the control group that was taught by human teachers without chatbots. Next, five studies compared the outcomes of two different chatbot types with different modalities or different instructional elements, such as a chatbot with a particular mechanism versus a chatbot without it (Hwang et al., 2022), or a chatbot with gamified elements versus a chatbot without gamified elements (Chien et al., 2022). Only one study compared the effects of interacting with chatbots with those of using traditional methods, such as CD players or paper-based materials (Hsu, Chen, & Yu, 2021). Last, in three studies, the effects of chatbots with more than two different modalities (or instructional elements) with those of human partners were compared. For example, Xu et al. (2022) compared the effects on children's reading comprehension and engagement of chatbots with or without dialogic strategies and human partners with or without dialogic strategies.

3.2.3. Location of chatbot use

As shown in Table 5, participants used chatbots in different contexts. The most frequently used location was a laboratory setting ($n = 13$) (e.g. Xu et al., 2022). In 10 studies, researchers employed chatbots in regular school language classes (e.g. Jeon, 2022b). Six studies examined the use of chatbots outside the classroom to supplement formal language instruction or the informal use of chatbots in naturalistic settings (e.g. Jeon, 2022a), and three had participants employ chatbots whenever appropriate both inside and outside the classroom (e.g. Wang, Pang, et al., 2022).

3.2.4. Participant information, intervention duration, and target language

Table 5 also displays distributions of the number of participants in the selected studies and their educational levels. Regarding the latter, higher education settings were given the most attention ($n = 17$), followed by elementary school ($n = 8$), secondary school ($n = 6$), and preschool settings ($n = 2$), while only one research study involved working adults. The number of participants in the selected studies ranged from 10 to 327. Six studies had fewer than 30 participants, nine studies from 30 to 60, and 17 studies more than 60 participants. Concerning intervention duration,12 studies involving only one to three learner-chatbot interactions were implemented in less than one week, 14 studies lasted more than one month, and six had durations ranging from one week to one month. Most of the studies ($n = 29$) had English as a second or foreign language (i.e. L2 English) as the target language.

3.3. Technological aspects

3.3.1. Types of adopted devices

Figure 3 shows seven device types used for chatbots in the selected studies, among which smartphones ($n = 8$), tablet PCs ($n = 8$), and smart speakers without a screen display ($n = 8$) were most used. Next, desktop PCs and smart speakers with a screen display were used in seven and five studies, respectively. In most studies, researchers used one device type, but in a few cases, they employed two or more types, comparing their effectiveness. For example, Tai and Chen (2022a, 2022b) employed chatbots installed on smart speakers either with or without a screen display and compared the effectiveness and learner perceptions of the two device types.

3.3.2. Roles of chatbot technology in language learning

As presented in Figure 4, we examined the roles that chatbots assumed in research treatment contexts and identified five: conversational partner, feedback provider, resource provider, evaluator, and interviewer. Conversational partner ($n = 26$) was found to be the most prevalent chatbot role

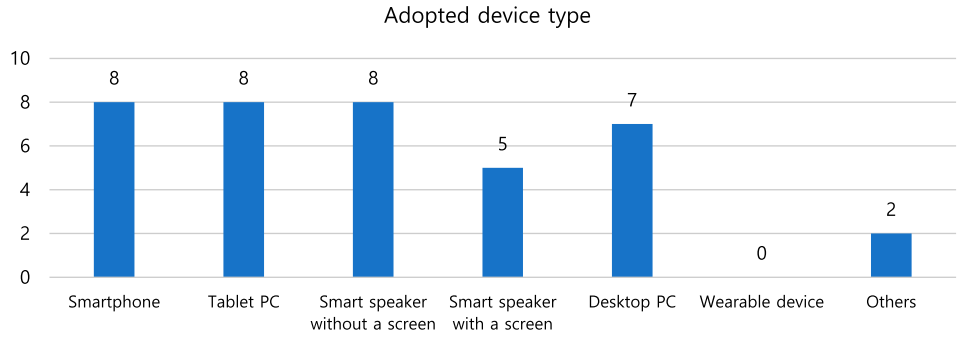

Figure 3. Adopted device types.

reported in the reviewed studies, followed by feedback provider ($n = 12$), resource provider ($n = 7$), evaluator ($n = 4$), and interviewer ($n = 1$). In most studies, chatbots performed one particular role throughout the treatment, such as conversational partner (e.g. Gonulal, 2021). However, in a few cases, some chatbots took more than one role. For example, in Wang, Liu, et al. (2022), the chatbot assumed the roles of feedback provider, resource provider, and evaluator.

4. Discussion

4.1. The changing landscape of research on speech-recognition chatbots

We found an upward trend in the quantity of published research on speech-recognition chatbots for language learning in recent years. This trend started in 2020 and accelerated at an exponential rate across 2021 and 2022, which differs from Bibauw et al.'s (2019) identification of 2007 as the year of the upswing in publications of research on both text- and speech-based chatbots for language learning. The more recent as well as more pronounced upward trend may reflect differences in the databases from which articles were selected or more likely differences in the chatbots on which studies in the two reviews were focused. Our focus was on chatbots with speech-recognition functionality (Shadiev & Liu, 2023), a highly accessible technology which was not widely available to researchers at earlier stages of chatbot development. Specifically, as reported in the two initial studies in our data, earlier researchers used chatbot systems that were not open to users (Morton & Jack, 2010) or could not be easily replicated by researchers who lacked sufficient programming knowledge (Hassani et al., 2016).

With the increasing advances in AI, speech-recognition chatbots have become more accessible for use and development, as is evident in recent studies. Additionally, we anticipate that the emergence of LLM-powered chatbots will further accelerate the utilization of these technologies. In this review, we identified two emerging trends in the selected studies. First, as mentioned in Huang et al. (2022) and Jeon (2022b), the emergence of visual chatbot development platforms has made it possible for researchers to develop their own chatbots even without advanced programming knowledge. For example, many researchers who published articles in 2021 and 2022 used such platforms to create their own chatbots, for example, Dialogflow (e.g. Hwang et al., 2022; Jeon, 2022b; Lee & Jeon, 2022; Xu et al., 2021, 2022).

Second, in contrast with previous studies, beginning in 2020, much research attention has been given to commercial speech-recognition chatbots that can be accessed by any user. For example, researchers took advantage of the increased prevalence of IPAs to investigate the uses of chatbots as pedagogical tools in language classes, for such purposes as providing target language conversational partners or delivering customized language learning resources such as quizzes and read-aloud stories. Some researchers found affordances on interaction

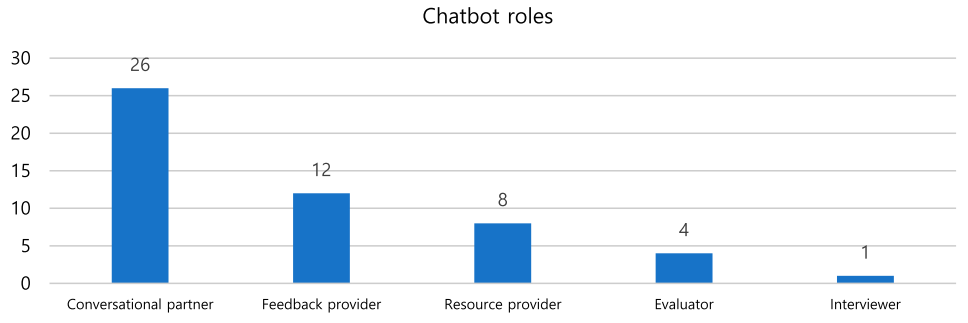

Figure 4. Chatbot roles.

capability lists of IPA manufacturers, such as Google's can-do-list or Amazon's Alexa skills kit (e.g. Dizon, 2020; Gonulal, 2021; Tai, 2022; Tai & Chen, 2022a, 2022b). Others examined the effects of existing speech-recognition chatbots that were designed for language learning (e.g. Jeon, 2022a; Wang, Liu, et al., 2022).

These most recent studies sharply contrast with earlier studies, such as Hassani et al. (2016) or Morton and Jack (2010), in which the complex development procedures for creating chatbots were not replicable by researchers without requisite programming knowledge. Also, at that time, researchers were challenged to find existing chatbots with speech-recognition functionality. Due to their enhanced accessibility to researchers and practioners, the emergence of LLM-powered chatbots is expected to accelerate the adoption of speech-recognition chatbots, surpassing even IPAs. Consequently, researchers may be expected to actively explore ways of utilizing these chatbots in pedagogically meaningful ways, tailored to particular educational contexts. It is also conceivable for researchers to receive technical assistance from these chatbots in constructing their own customized chatbots that align with their investigative and educational objectives (Jeon, 2022b).

4.2. The need for more diversified approaches to conducting speech-recognition chatbot research

As demonstrated by the increasing societal interest in LLM-powered chatbots, chatbot technology is poised to become increasingly integrated into education. This integration necessitates rigorous and balanced research efforts, underpinned by a comprehensive understanding of the existing relevant literature. In this review, we found that, while more studies have confirmed the positive effects of using speech-recognition chatbots on language learning, more diversified approaches to researching speech-recognition chatbots in much broader contexts are needed. Regarding research focus, methodology, and target language, we discovered that in most studies a mixed-methods approach was adopted to examine how learners of English perceived chatbots or advanced in their language proficiency after using them. These findings suggest that future researchers, including those exploring the educational uses of LLM-powered chatbots with speech-recognition functionality, should focus on aspects that have received less attention, such as cultural factors, writing and reading as well as oral language skills, learner behaviors, and individual learner characteristics. We also recommend that researchers adopt varied research approaches, including advanced quantitative or in-depth qualitative methods, and that they target learners of languages other than English.

Regarding group design, single-chatbot group designs were adopted in many studies. Recently, more studies have demonstrated the positive effects of using chatbots for language learning by comparing outcomes from using and not using them (e.g. Jeon, 2022a; Tai, 2022). Having established the value of this technology, researchers now need to bolster evidence of the effectiveness of specific features of chatbots and of various pedagogical approaches to using them. The need for

this research will increase as the implementation of various chatbots such as OpenAI's GPT series, Google's Bard, and Meta's BlenderBot gains increasing prominence. These chatbots, moreover, can be incorporated with different platforms that possess distinct features. For example, chatbots in the GPT series are now integrated within web search engines, allowing them to serve as co-pilots for users as they navigate and learn from the vast expanse of the internet. When receiving questions, Google's Bard generates multiple drafts of a response concurrently, enabling users to critically evaluate and select the most appropriate answer. Meta's BlenderBot platform incorporates human-like features such as visual elements that simulate eyes and persona-specific capabilities. In addition, ChatGPT and GPT-4 deliver responses as if they were talking in real-time, while Bard and BlenderBot present completed answers. As technological advancements continue to diversify the features of chatbot platforms, a broader range of educational affordances will emerge, further enriching the learning experience for students. The emergence of novel affordances offered by these chatbots will continue to warrant rigorous scholarly investigation (Jeon, 2022b; Jeon et al., 2022). In this regard, we recommend that future researchers focus more on comparisons among different types of chatbots with different functions and among uses of different pedagogical strategies with the same chatbot (Lin & Mubarok, 2021). On a similar note, most studies we reviewed were conducted in lab or classroom settings. As chatbots are ubiquitous in nature, more naturalistic investigations of ways of using chatbots for learning outside the classroom are needed to establish the ecological validity of such uses (Jeon, 2022a).

Last, we found the need for more research on different groups of participants and for more long-term studies. Similar to findings from other reviews of studies on chatbots (e.g. Ji et al., 2023), we found that higher education was the most preferred research setting. Few studies were conducted with working adults or preschoolers. Further research involving more diversified educational levels, with particular attention to working adults and younger learners as under-researched groups, is needed. Regarding intervention duration, only around half of the reviewed studies were conducted for more than one month, and in the majority of those conducted within a one-month period, students were exposed to a small number of chatbot tasks. As indicated by Huang et al. (2022) and Fryer et al. (2020), it is important to consider the novelty effect when introducing chatbots into the learning context. Thus, we recommend that future studies be designed to extend over longer periods or that measures be taken to minimize the novelty effect before implementing short-term interventions.

4.3. Toward broader ubiquity and greater sense of immersion

We found that many speech-recognition chatbots were used on mobile devices, including smartphones, tablet PCs, and smart speakers, rather than traditional desktop PCs. This portability enables learners to have unlimited access to chatbots, but we found no research on the use of chatbots with wearable devices such as smart watches, which are the most portable and potentially the most ubiquitous of mobile devices (Bower & Sturman, 2015). On the other hand, instead of pursuing the breadth of ubiquity, some researchers have pursued the depth of immersion into environments in which realistic interactions between chatbots and students can occur by incorporating other state-of-art technologies. For example, Divekar et al. (2021) used several virtual reality (VR)- and augmented reality (AR)-related technologies featuring panoramic settings and equipped with various sensors, wireless lapel microphones, and ambient sound systems to investigate how such immersive environments can support language learning. These two trends toward broader ubiquity and deeper immersion are gaining further momentum as attempts have been made to incorporate LLM-powered chatbots into different types of technologies, including mobile interfaces, computer vision technology, such as image generation and recognition, and the metaverse (Kasneci et al., 2023; OpenAI, 2023; Zhang et al., 2023).

Therefore, we argue that, in addition to applications using conventional devices, researchers need to investigate the use of two other technological approaches to using speech-recognition chatbots.

One is the integration of chatbots into wearable devices, which may provide novel affordances, interaction experiences, and learning effects as well as maximize chatbots' potential for ubiquity (Jeon, 2022b). For example, researchers could explore the integratation of LLM-powered chatbots into mobile applications for wearable devices by utilizing the APIs of ChatGPT or GPT-4 to develop more ubiquitous LLM-powered chatbots with speech-recognition functionality. The other is to combine chatbots with other advanced technologies, such as VR or AR-related devices with head-mounted devices or panoramic screens, to create an experience of immersion in an authentic environment for learner-chatbot interaction. Furthermore, as LLM-powered chatbots evolve to integrate computer vision technology, it becomes increasingly pertinent to examine the impact of multimodal interactions between chatbots and humans on aspects of the learning process, such as learner motivation, engagement, and immersion. In addition, Rospigliosi (2023) posits that LLM-powered chatbots will be incorporated into emerging virtual environments such as the metaverse, further highlighting the need for comprehensive research on educational chatbots. However, at present, how such environments can be designed with what specific technologies and with what effects on learning remains underexplored.

4.4. Expanding the chatbot's role beyond that of conversational partner

Despite chatbots' potential for assuming various roles in the learning process (Ji et al., 2023), in the studies we reviewed, chatbots mostly performed the role of conversational partner. This situation contrasts with studies of text-based chatbots, in which, besides serving as conversational partners, they have performed such functions as analyzing students' needs (Jeon, 2021), fostering students' goal-setting and social presence (Hew et al., 2023), and providing online video guidance (Fidan et al., 2022). We recommend that future researchers similarly involve speech-recognition chatbots in different roles such as those in studies of text-based chatbots and examine the effects of these roles on learning in both modalities. In addition, while LLM-powered chatbots offer vast potential for fulfilling diverse roles in educational contexts, their specific applications have yet to be thoroughly investigated. As a result, identifying suitable pedagogical uses for these chatbots and determining the roles they can play as either self-directed learning tools or assistants for human educators present important directions for future research.

5. Limitations

Due to some limitations, the findings of this systematic review of selected studies should be interpreted with caution. First, we employed a thematic approach. While this approach served our research purpose of providing a comprehensive analysis of relevant studies (e.g. Hwang & Chang, 2021; Huang et al., 2022; Liang et al., 2021; Shadiev et al., 2017), it is limited by its descriptive nature. To further advance knowledge of chatbots with speech-recognition functionality, future researchers might employ more in-depth analytic techniques, such as bibliographic, sentiment, or meta-analysis (Bibauw et al., 2022). Second, as studies are now often published as early-access articles, minor changes to publication years are possible as articles are assigned to issues. To assist readers in understanding the publication trend more accurately, we provided the search date in Figure 1. Last, the process of conducting a literature search does not guarantee the inclusion of all previous studies on speech-recognition chatbots. It may be necessary for future researchers to consult other databases and use additional search terms.

6. Implications, suggestions, and conclusion

With the rapid development of AI technologies such as ASR and NLP and the increasing attention to AI in education, researchers have recently become more interested in the potential of speech-recognition chatbots to enhance students' learning. As human-computer interaction has become an established practice in the era of LLMs, the uses, roles, and ranges of chatbots are expected to

continue to expand exponentially. Given this societal change and an expanding body of research on speech-recognition chatbots, we conducted a systematic review of relevant publications to identify research trends and suggest future research directions.

This study has important implications for the field of educational technology in general and chatbot-assisted language learning in particular. Rather than addressing text- and speech-based chatbots as one group, we specifically explored research on speech-recognition chatbots as an emerging area of inquiry. With this focus, we discovered factors related to research on chatbots with speech-recognition functionality that were not apparent in previous reviews (e.g. Bibauw et al., 2019; Huang et al., 2022; Ji et al., 2023). These factors included the predominant focus on speaking and listening skills, the different dating of the upward trend in numbers of relevant publications, and device types commonly used for accessing chatbots. Also, we extracted aspects that were not considered in previous reviews but have recently become the focus in many empirical studies (e.g. Fryer et al., 2020; Ji et al., 2023). These include group design formats and settings in which chatbots are used. In addition, based on several matrixes grounded in previous review studies (e.g. Elaish et al., 2022; Shadiev et al., 2017), we have presented a comprehensive analysis of speech-recognition chatbot research. This study may thus help researchers and practitioners gain insight into trends in research on speech-recognition chatbots and future directions of their applications in the era of LLM-powered chatbots.

Based on our analysis, we provide four suggestions for researchers investigating speech-recognition chatbots in language learning and adjacent fields. First, researchers seeking to explore their educational possibilities can use either existing commercial chatbots such as LLM-powered chatbots or IPAs (e.g., Tai & Chen, 2022a) or develop their own speech-recognition chatbots using visual chatbot development platforms such as Dialogflow (Jeon, 2022b; Lee & Jeon, 2022). Second, researchers might adopt more diversified approaches to investigating speech-recognition chatbots. In this regard, we suggest that researchers consider the eight research design aspects we used in this study (i.e. research focus, methodology, group design format, setting of chatbot use, numbers of participants, participants' educational level, intervention duration, and target language). We also recommend that they focus on aspects found to be underexplored in the previous literature. Third, we propose that researchers expand their focus beyond chatbots installed on traditional mobile devices to their adaptability to two types of new technologies: wearable technology (e.g., smartwatches) and multimodal technology (e.g., VR/AR applications, computer vision, and metaverse). Their installation on wearable devices can broaden chatbots' natural propensity for ubiquity. Also, coupling chatbots with the metaverse or various stage technologies (e.g., additional devices for VR and AR, panoramic screens, or ambient sound technologies) may enhance the sense of immersion that chatbots can provide (Divekar et al., 2021; Rospigliosi, 2023). Exploring the effects of multimodal interactions between students and chatbots, including images, videos, and gestures, on the learning process also constitutes a compelling avenue for future research. Providing affordances associated with novel technologies may result in novel learner-chatbot interaction experiences and learning outcomes (Jeon, 2022b). Last, researchers might investigate and develop roles for speech-recognition chatbots beyond that of conversational partner. As suggested eariler, researchers can examine how the adaption of various roles currently assumed by text-based chatbots (Hew et al., 2023; Jeon, 2021) to speech-recognition modality may affect learners' interaction experiences, perceptions, and learning outcomes. Similarly, additional research is needed to explore the full range of roles that LLM-powered chatbots can play in educational settings and to compare their educational possibilities with and without speech-recognition capabilities.

In conclusion, this focused systematic review contributes to the fields of technology-enhanced language learning and chatbot research by providing novel insights specific to studies on speech-recognition chatbots. This work may serve as a valuable reference for researchers, developers, and practitioners seeking to understand, advance, and utilize speech-recognition chatbots, including various types of LLM-powered chatbots, as educational tools that can effectively facilitate learning.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jaeho Jeon is a doctoral student in the department of Literacy, Culture, and Language Education at Indiana University Bloomington. His research interests include computer-assisted language learning, dynamic assessment, and teacher education. His articles have been published in *Computer Assisted Language Learning*, *Computers & Education*, *Interactive Learning Environments, Education and Information Technologies*, *System*, among others.

Seongyong Lee currently serves as an assistant professor in the department of English Education at Hannam University, South Korea. He obtained his PhD in second and foreign language education from the State University of New York at Buffalo. His research interests include computer-assisted language learning, world Englishes, English as a lingua franca (ELF) in academic settings, and English as a second language writing. His articles have appeared in *Computer Assisted Language Learning, Computers & Education, Information Technology & People*, *English Today*, *Asian Englishes*, *Applied Linguistics Review, Lingua, System*, among others.

Seongyune Choi is a Ph.D. candidate in the Department of Computer Science and Engineering at Korea University in Seoul, South Korea. His research interests include computer science education, educational technologies, and artificial intelligence in education. His articles have appeared in *International Journal of Human-Computer Interaction*, *Education and Information Technologies*, *Educational Technology & Society*, among others.

ORCID

Jaeho Jeon http://orcid.org/0000-0002-1161-3676
Seongyong Lee http://orcid.org/0000-0002-9436-4272
Seongyune Choi http://orcid.org/0000-0002-7968-3726

References

*Indicates the studies selected for review.

*Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward personalized scaffolding and fading of motivational support in L2 learner–dialogue agent interactions: An exploratory study. *IEEE Transactions on Learning Technologies*, *13*(3), 604–616. https://doi.org/10.1109/TLT.2020.2989776

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, *32*(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

Bibauw, S., François, T., Noortgate, V. d., & Desmet, W. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology*, *26*(1), 1–24. https://doi.org/10125/73488

Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computers & Education*, *88*, 343–353. https://doi.org/10.1016/j.compedu.2015.07.013

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

*Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2020.1833043

*Chien, Y., Wu, T., Lai, C., Huang, Y. (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

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk*, *34*(5), 545–567. https://doi.org/10.1515/text-2014-0018

Davis, R. O., Vincent, J., & Park, T. (2019). Reconsidering the voice principle with non-native language speakers. *Computers & Education*, *140*, 103605. https://doi.org/10.1016/j.compedu.2019.103605

*Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2021). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2021.1879162

*Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, *24*(1), 16–26. https://doi.org/10125/44705

*Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2022.2096638

Elaish, M. M., Hussein, M. H., & Hwang, G. (2022). Critical research trends of mobile technology-supported English language learning: A review of the top 100 highly cited articles. *Education and Information Technologies*. Advance online publication. https://doi.org/10.1007/s10639-022-11352-6

Fidan, M., & Gencel, N. (2022). Supporting the Instructional Videos With Chatbot and Peer Feedback Mechanisms in Online Learning: The Effects on Learning Performance and Intrinsic Motivation. *Journal of Educational Computing Research*, *60*(*7*), 1716–1741. http://dx.doi.org/10.1177/07356331221077901

*Forsyth, C. M., Luce, C., Zapata-Rivera, D., Jackson, G. T., Evanini, K., & So, Y. (2019). Evaluating English language learners' conversations: Man vs. machine. *Computer Assisted Language Learning*, *32*(4), 398–417. https://doi.org/10.1080/09588221.2018.1517126

*Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

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

*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

Fu, Q., & Hwang, G. (2018). Trends in mobile technology-supported collaborative learning: A systematic review of journal publications from 2007 to 2016. *Computers & Education*, *119*, 129–143. https://doi.org/10.1016/j.compedu.2018.01.004

*Gonulal, T. (2021). Investigating EFL learners' humorous interactions with an intelligent personal assistant. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.1974489

*Hassani, K., Nahvi, A., & Ahmadi, A. (2016). Design and implementation of an intelligent virtual environment for improving speaking and listening skills. *Interactive Learning Environments*, *24*(1), 252–271. https://doi.org/10.1080/10494820.2013.846265

Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education*, *35*(1), 40–68. Advance online publication. https://doi.org/10.1007/s12528-022-09338-x

*Hsu, H. L., Chen, H. H. J., & Todd, A. G. (2021). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.2016864

Hsu, L. (2022). To CALL or not to CALL: Empirical evidence from neuroscience. *Computer Assisted Language Learning*, *35*(4), 792–815. https://doi.org/10.1080/09588221.2020.1750429

*Hsu, M., Chen, P., & Yu, C. (2021). Proposing a task-oriented Chatbot system for EFL learners speaking practice. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.1960864

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, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.1952615

Hwang, G.-J., & Tsai, C.-C.. (2011). Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, *42*(*4*), E65–E70. http://dx.doi.org/10.1111/bjet.2011.42.issue-4

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

Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2021.1987272

*Jeon, J. (2022a). Exploring a self-directed interactive app for informal EFL learning: A self-determination theory perspective. *Education and Information Technologies*, *27*(4), 5767–5787. https://doi.org/10.1007/s10639-021-10839-y

*Jeon, J. (2022b). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2021.2021241

Jeon, J, Lee, S, & Choe, H. (2022). Enhancing EFL pre-service teachers' affordance noticing and utilizing with the Synthesis of Qualitative Evidence strategies: An exploratory study of a customizable virtual environment platform. *Computers & Education*, *190*, 104620. http://dx.doi.org/10.1016/j.compedu.2022.104620

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, *55*(1), 48–63. Advance online publication. https://doi.org/10.1080/15391523.2022.2142873

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. http://dx.doi.org/10.1016/j.lindif.2023.102274

*Lee, S., & Jeon, J. (2022). Visualizing a disembodied agent: Young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2067182

Liang, J., Hwang, G., Chen, M. A., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2021.1958348

*Lin, C., & 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.

Moher, D., Liberati, A., Terzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

*Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: A cross-cultural evaluation. *Computer Assisted Language Learning*, *23*(4), 295–319. https://doi.org/10.1080/09588221.2010.493524

*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

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220

OpenAI. (2023, March 31). *GPT-4.* https://openai.com/research/gpt-4.

Rospigliosi, P. A.. (2023). Artificial intelligence in teaching and learning: what questions should we ask of ChatGPT?. *Interactive Learning Environments*, *31*(*1*), 1–3. http://dx.doi.org/10.1080/10494820.2023.2180191

Shadiev, R., Hwang, W., & Huang, Y. (2017). Review of research on mobile language learning in authentic environments. *Computer Assisted Language Learning*, *30*(3-4), 284–303. https://doi.org/10.1080/09588221.2017.1308383

Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, *35*(1), 74–88. Advance online publication. https://doi.org/10.1017/S095834402200012X

*Tai, T. Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2075013

*Tai, T. Y., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2020.1841801

*Tai, T. Y., & Chen, H. H. J. (2022a). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2040536

*Tai, T. Y., & Chen, H. H. J. (2022b). The impact of intelligent personal assistants on adolescent EFL learners' listening comprehension. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2040536

*Timpe-Laughlin, V., & Dombi, J. (2020). Exploring L2 learners' request behavior in a multi-turn conversation with a fully automated agent. *Intercultural Pragmatics*, *17*(2), 221–257. https://doi.org/10.1515/ip-2020-0010

Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2022). Using spoken dialogue technology for L2 speaking practice: What do teachers think? *Computer Assisted Language Learning*, *35*(5-6), 1194–1217. https://doi.org/10.1080/09588221.2020.1774904

*Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2032184

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, *29*(4), 833–851. https://doi.org/10.1080/09588221.2016.1167090

*Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2022). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*. Advance online publication https://doi.org/10.1016/j.compedu.2022.104703.

*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*. Advance online publication. 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*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Xie, H., Chu, H., Hwang, G., & Wang, C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Xu, W., & Zammit, K. (2020). Applying thematic analysis to education: A hybrid approach to interpreting data in practitioner research. *International Journal of Qualitative Methods*, *19*, 1–9. https://doi.org/10.1177/1609406920918810

*Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim Y., & Warschauer, M. (2022). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development*, *93*(2), 149–167. https://doi.org/10.1111/cdev.13708

*Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*, 104059. https://doi.org/10.1016/j.compedu.2020.104059

*Yang, C. T., Lai, S., & Chen, H. H. (2022). The impact of intelligent personal assistants on learners' autonomous learning of second language listening and speaking. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2022.2141266

*Yang, H., Kim, H., Lee, J., & 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

Zhang, Y. (2023, March). *Universal Speech Model (USM): State-of-the-art speech AI for 100+ languages*. Google Research. https://ai.googleblog.com/2023/03/universal-speech-model-usm-state-of-art.html.