

79. Chatbots in English as a foreign or second language education contexts: A review of recent empirical research

Hatice OKYAR¹

APA: Okyar, H. (2023). Chatbots in English as a foreign or second language education contexts: A review of recent empirical research. *RumeliDE Dil ve Edebiyat Arařtırmaları Dergisi*, (36), 1333-1346. DOI: 10.29000/rumelide.1369170.

Abstract

Chatbots, as one of the conversational applications of artificial intelligence technology, have been the focus of attention in the field of second/foreign language (L2) acquisition research. However, very few studies have so far reviewed the literature on chatbots and L2 learning, in particular the learning of English as a foreign/second language (EFL/ESL). Based on this, this paper set out to review studies that are related to the use of chatbots in EFL/ESL learning contexts. The study included eleven empirical articles published in the Social Science Citation Index (SSCI) journals between the years 2018 and 2022, and all of them were analyzed in detail considering the following three aspects: 1) research features such as data collection methods, research method, and study context; 2) chatbots included in the studies; 3) results of the studies. The findings showed that all the studies were conducted in EFL contexts and most of them used a mixed-method design. It was identified that different types of chatbots were used in the studies, and the studies generally focused on improving learners' speaking skills. The findings also revealed that the use of chatbots generally contributed positively to the EFL learning process. Furthermore, the review identified that chatbot-integrated learning had some affective/psychological outcomes. In addition to the many positive effects brought by the use of chatbots, it was also identified that chatbots had some limitations. Finally, this review study made some suggestions for future research based on the findings of this study and the identified gaps in the literature.

Keywords: Chatbots, chatbots in EFL/ESL contexts, artificial intelligence, systematic review

Yabancı veya ikinci dil olarak İngilizce eğitim bağlamlarında sohbet robotları: Son zamanlardaki ampirik arařtırmaların incelenmesi

Öz

Yapay zeka teknolojisinin konuşma uygulamalarından biri olan sohbet robotları, ikinci/yabancı dil edinimi araştırma alanında ilgi odağı olmuştur. Bununla birlikte, şimdiye kadar çok az sayıda çalışma sohbet robotları ve yabancı dil öğrenimi, özellikle de İngilizcenin yabancı dil olarak öğrenimi üzerine olan literatürün incelemesini yapmıştır. Buna dayanarak, bu çalışma, sohbet robotlarının İngilizcenin yabancı/ikinci dil olarak öğrenildiği bağlamlarda kullanılması üzerine yapılmış olan çalışmaları incelemeyi amaçlamaktadır. Çalışma, 2018-2022 yılları arasında, Sosyal Bilimler Atf Dizinindeki dergilerde yayımlanmış olan 11 adet kanıta dayalı (ampirik) çalışmaları incelemiştir ve bu çalışmalar, şu üç madde göz önüne alınarak detaylı bir şekilde analiz edilmiştir: a) veri toplama yöntemi, araştırma deseni, çalışma bağlamı; b) çalışmalara dahil edilen sohbet robotları; c) çalışmaların sonuçları. Bulgular tüm çalışmaların, İngilizcenin yabancı dil olarak öğrenildiği

¹ Dr. Öğr. Üyesi, Necmettin Erbakan Üniversitesi, Yabancı Diller Yüksekokulu (Konya, Türkiye), okyarhatice@gmail.com, ORCID ID: 0000-0003-4458-4805 [Araştırma makalesi, Makale kayıt tarihi: 11.08.2023-kabul tarihi: 20.10.2023; DOI: 10.29000/rumelide.1369170]

bağlamlarda gerçekleştirildiğini ve çalışmaların çoğunun karma yöntem tasarımı kullandığını göstermiştir. Çalışmalarda farklı türden sohbet robotlarının kullanıldığı ve çalışmaların genellikle öğrencilerin konuşma becerisini geliştirmeye odaklandıkları belirlenmiştir. Bulgular ayrıca sohbet robotları kullanımının genel olarak yabancı dil olarak İngilizce öğrenme sürecine pozitif katkıda bulunduğunu göstermiştir. Bunlara ek olarak, inceleme, sohbet robotu entegreli öğrenmenin duyuşsal/psikolojik etkilerinin de olduğunu belirlemiştir. Sohbet robotu kullanımının getirdiği pozitif etkilere ek olarak, sohbet robotlarının bazı eksikliklerinin de olduğu belirlenmiştir. Son olarak, bu inceleme çalışması, çalışmanın bulgularına ve literatürdeki eksikliklere dayanarak gelecekteki çalışmalar için bazı önerilerde bulunmuştur.

Anahtar kelimeler: Sohbet robotları, yabancı/ikinci dil olarak İngilizce bağlamlarında sohbet robotları, yapay zeka, sistematik inceleme

1. Introduction

Artificial intelligence (AI)-based applications and tools, just like in other fields (e.g. digital assistance, medicine/healthcare, marketing, security), have started to become a significant part of education in recent years. One of the most popular of these AI tools is chatbots, which are defined as “a software tool that interacts with users on a certain topic or in a specific domain in a natural, conversational way using text and voice” (Smutny & Schreiberova, 2020, p. 1). Considering the fact that the first chatbot named ELIZA emerged in 1966 and right after different types of chatbots such as HeX, Alaude, and ALICE continued to be developed, it is possible to say that chatbot applications and related research have existed for a very long time (Smutny & Schreiberova, 2020; Yasmin & Mazhar, 2023). Equipped with natural language processing and speech recognition technology, AI-powered chatbots can simulate everyday spoken language (Yasmin & Mazhar, 2023), and this has made them an important educational actor in teaching and learning processes (Bhise et al., 2023). Deng and Yu (2023) stated that “chatbots perform three roles in education, i.e., teaching assistants, learning partners, and personal tutors.” (p. 2). These three functions are closely connected to each other, and Deng and Yu (2023) explain them as follows: When chatbots are evaluated as a teaching assistant, it can be said that they support students' learning on the online platform, supply them with necessary information, and allow them to receive immediate feedback. Additionally, when they are considered as a learning partner, they can communicate with learners both in written and spoken forms. As a tutorial partner, chatbots can provide learners with questions and answers, etc. (Deng & Yu, 2023).

There are many reasons behind the interest in using and integrating chatbots in education as follows: offering learners the opportunity to ask questions and get answers whenever and wherever they want, i.e., without the limit of time and space (Bhise et al., 2023; Fryer & Carpenter 2006); providing synchronous interaction (Huang et al., 2022); helping permanent learning by allowing unlimited repetition of information and content, in other words, helping learners retain information in their minds (Deng & Yu, 2023); reducing the workload of teachers (Perez et al., 2020); giving instant feedback to learners (Huang et al., 2022); arousing and maintaining motivation and interest for learning in students (Deng & Yu, 2023; Fryer & Carpenter 2006; Han, 2020; Han et al., 2022); leading to success in (language) learning (Deng & Yu, 2023; Lee & Hwang, 2022); improving speaking skill (Han, 2020); encouraging shy learners by providing the opportunity of anonymous interaction (Pereira et al., 2019); and giving opportunity for self-paced learning (Deng & Yu, 2023).

Chatbots, as in many fields of education, have aroused the interest of researchers and educators in the field of second language acquisition because they provide learners with communication and interaction opportunities in L2, and thus, help them learn and practice daily language (e.g. Han, 2020; Jeon, 2022; Lee & Hwang, 2022). According to Fryer and Carpenter (2006), there are six advantages that chatbots offer to language learners: 1. Learners feel more comfortable talking with chatbots rather than talking with humans, 2. The same learning topic can be repeated over and over by chatbots, and they do not complain or get annoyed while doing this, 3. Learners can improve their reading and listening skills as chatbots can offer text and speech input, 4. Many learners find learning with chatbots very interesting, 5. Using chatbots, learners can encounter many different vocabularies and structures of the target language, 6. Learners can get instant feedback from chatbots for their grammar and spelling mistakes (Fryer & Carpenter, 2006). Also, as for English as a foreign language (EFL) learning and teaching contexts, various empirical research studies have proved the positive impacts of chatbot use (Jiang, 2022). These results are promising because, especially in foreign language learning contexts, students do not have the opportunity to practice the target language enough or to get sufficient feedback on their use of the language in their daily lives or at school (Fryer & Carpenter, 2006). For that reason, to improve the process of foreign language learning and to make this process more effective, the number of studies applying technological advances to language learning contexts has begun to increase. Chatbots are one of them. On this basis, it is important to have a collective view and evaluation of studies on the use of chatbots in foreign language education. Although review and meta-analysis studies on the use of chatbots in the field of general education are more common (e.g. Deng & Yu, 2023; Hwang & Chang, 2021; Perez et al., 2020), there are a limited number of systematic reviews specific to the role of chatbots in the field of foreign language education. For instance, Huang et al. (2022) made a systematic review of 25 studies that included articles related to first, second, and foreign language learning and published until the end of October 2020. They found that using chatbots has technological facilities such as timeliness, (e.g. getting instant feedback, synchronous interaction), ease of use (e.g. being easily accessible), and personalization (e.g. chatbots make it possible for each learner to have a unique experience). They also identified five pedagogical uses of chatbots such as a) being an interlocutor (e.g. accompanying and helping learners during their learning process), b) being a simulation of a real language learning setting through activities such as role-plays, c) information transfer i.e., conveying the learning content to learners, d) being a helpline (Students can get answers as soon as they submit their questions), e) giving recommendations related to the learning needs of learners such as books. Additionally, Huang et al. (2022) found that chatbots also provide learners with social affordances such as interpersonal (self-disclosure of learners by sharing information about themselves), open (e.g. expression of agreement), and cohesion communication (e.g. use of greetings). Another research, conducted by Zhai and Wibowo (2022), systematically reviewed 48 articles published between 2012 and 2022 to investigate how the cultural, empathetic, and humorous dimensions of chatbots influence L2 learners' learning. Their findings indicated that these dimensions can have positive effects on the use of chatbots for improving learners' L2 learning. Unlike Huang et al.'s (2022) and Zhai and Wibowo's (2022) studies, Lee and Hwang (2022) focused on a particular context and conducted a meta-analysis of 16 experimental studies on the use of chatbots in the Korean EFL context. Their findings showed that chatbots can be used to support and improve the language skills of learners.

The contributions of these studies to the field are invaluable, however, the scarcity of such research clearly indicates the need for more review studies with different perspectives on the applications of chatbots in L2 contexts, especially in EFL/ESL contexts. As previously mentioned, with significant advances in artificial intelligence technology and new developments related to the use of chatbots in the foreign language learning field, the number of empirical studies exploring the integration of chatbots in

the EFL/ESL learning domain is increasing day by day. Therefore, it is important to keep reviews on this topic up to date to identify recent research applications and trends. Conducting updated reviews that examine the studies from different aspects will help researchers to be aware of the developments and gaps in the existing literature and finally will help them improve it with better-designed research in the future. Considering these important reasons, the current systematic review aims to examine the empirical studies published between 2018-2022 on the use of chatbots in the field of EFL/ESL, and to contribute to the existing chatbot-based second language learning research by focusing on the following three main issues: 1) research characteristics; 2) types of chatbots; 3) findings of the studies. In line with this purpose, the present review is guided by the research questions below:

1. What are the research features of the studies in terms of research methods, data collection tools, education, and study context?
2. What are the chatbots used in the studies?
3. What are the findings of the studies related to the use of chatbots in EFL/ESL contexts?

2. Methodology

The main focus of this systematic review study was to provide a detailed review of empirical articles on the use of chatbots in the context of EFL/ESL learning. To realize this aim, a topic search that included keywords, title, and abstract was conducted using the Web of Science (WoS) database, and the following search string was used by including Boolean operators as follows: (chatbot) AND (English OR English as a second language OR English as a foreign language OR English language learner OR ESL OR EFL). The publication date of the studies in-print or online was set from 1st of January 2018 to 31st of December 2022. In accordance with the research questions, the studies were evaluated using the inclusion and exclusion criteria given in Table 1:

Table 2. Article selection criteria

Inclusion Criteria

1. The study must include a chatbot and its use in the field of EFL/ESL education.
2. The study must present empirical data/results regarding the use of chatbots in the field of EFL/ESL education.
3. The study must include EFL/ESL learners as participants.
4. One of the aims or research questions of the study must be related to the effects of chatbot use in the field of EFL/ESL.
5. The study must provide information about the included chatbot.
6. The article must be published in SSCI-indexed journals between 2018 and 2022.

Exclusion Criteria

1. Studies written in languages other than English were excluded.
2. Proceeding papers, review articles, and meta-analysis studies were excluded from this review.
3. Studies that only report the design and development process of a chatbot; only report a pilot study of a chatbot, or only describe chatbots were excluded.
4. Studies not related to the field of EFL/ESL education were excluded.

Researchers such as Chu et al. (2022), and Xia and Zhong (2018) highlighted the importance of considering the quality of publications to be included in review studies. On the basis of this recommendation, the present review only included SSCI journal papers; therefore, it can be assumed that the studies in this review are of a good quality (Chu et al., 2022). The initial search process in the WoS database yielded 97 studies. After filtering the studies as only *articles* and language as only *English*,

and index as only *SSCI*, the number of articles obtained was 22. These remaining 22 articles were assessed based on the pre-determined inclusion/exclusion criteria and research questions. As a part of the first screening process, titles and abstract of each study were examined, and irrelevant studies were excluded. In the second process, the full text of the remaining articles was read thoroughly to verify that they all met the predefined inclusion criteria. Finally, a total of 11 articles that met the criteria were decided to be included in the systematic review. Although the time period for the articles was set as 2018-2022, the distribution of the included articles by years is as follows: 5 studies in 2021 and 6 studies in 2022. No study was found that met the inclusion criteria in the years 2018, 2019, 2020. During the analysis of the articles, a data-coding form was used to extract all relevant data from each study. The following main factors were examined in each article: a) research features such as research methods, data collection tools, education context/level, and study context/country; b) chatbots used in the studies; and c) results of the reviewed studies. Two weeks after the coding process was completed, the researcher re-read the content of all the articles and compared each with the data in the coding table to ensure that all relevant data were fully transferred to the final data sheet. The findings obtained are presented in the next section of the paper.

3. Findings

3.1. Findings related to the research features of the studies

The research methods as indicated and defined in the related studies are presented in Table 2. As seen in the table, of the 11 studies, 7 employed mixed methods, reporting findings of both qualitative and quantitative data. While 3 studies used quantitative methods, only 1 study in this review adopted a qualitative method to examine chatbot use in EFL/ESL contexts.

Table 2. Research methods

| Research methods | N |
|------------------|---|
| Mixed methods | 7 |
| Quantitative | 3 |
| Qualitative | 1 |

Various data collection methods were used in the studies examined. Among the quantitative data collection techniques, questionnaires/scales (n=11) such as motivation questionnaires (n=2); perception questionnaires (n=2); situational interest questionnaire (n=1); godspeed questionnaire (n=1); flow questionnaire (n=1); self-report questionnaire (n=1); social presence scale (n=1); human-likeness scale (n=1); and satisfaction survey (n=1) were the most commonly employed data collection tools, followed by pre- and post-tests (n=6). The qualitative data collection tools included interaction data/logs/records (n=6), interviews (n=5), and open-ended surveys (n=1). All of the studies adopted more than one data collection tool in their research. Also, the studies with mixed method design used both qualitative and quantitative data collection techniques, which allowed them to produce versatile, more comprehensive, and detailed results. The details of the data collection tools are presented in Table 3.

Table 3. Data collection tools

| Data collection tools | N |
|-----------------------|----|
| Questionnaire | 11 |
| Pre- and post-test | 6 |

| | |
|------------------------------|---|
| Interaction data/logs/record | 6 |
| Interview | 5 |
| Open-ended survey | 1 |

As for the study context, it was identified that all of the studies were conducted with EFL learners in EFL settings. The distribution of studies according to education context/level is as in the following: a) higher education (university) (n=5), b) primary/elementary schools (n=4), c) high school (n=1), and d) mixed (n=1) As Yang et al. (2022) conducted their study with two education levels, namely, elementary and high school, their study was coded in the mixed category in the table. Additionally, the names of the countries where the studies were conducted are as follows: a) Taiwan (n=4); b) Korea (n=3); c) Egypt (n=1); d) Iran (n=1). Hew et al. (2022) mentioned their study context as Asia. Also, one study did not clearly mention the name of the study country/region although it included EFL learners as participants, and so it was expressed as *not mentioned* in the table. Tables 4 and 5 present the distribution of studies by country and education context.

Table 4. Education context/level

| Education context/level | N |
|----------------------------|---|
| Higher education | 5 |
| Primary/elementary schools | 4 |
| High school | 1 |
| Mixed | 1 |

Table 5. Study context/country/region

| Study context/country/region | N |
|------------------------------|---|
| EFL/Taiwan | 4 |
| EFL/Korea | 3 |
| EFL/Egypt | 1 |
| EFL/Iran | 1 |
| EFL/Asia | 1 |
| EFL/Not mentioned | 1 |

3.2. Chatbots used in the studies

It was identified that studies generally used different types of chatbots with various features for EFL learners. Table 6 shows the names of the chatbots used in the studies.

Table 6. Chatbots used in the studies

| Used chatbots | Authors |
|--|-------------------------------------|
| Voice chatbot "Ellie" | Yang et al. (2022) |
| Mondly and web chatbots (Mike, Charles, Audrey, and Cristal) | El Shazly (2021) |
| LINE ChatBot | Chien et al. (2022) |
| SmartU English | Hwang et al. (2022) |
| AI chatbot built by Google Dialogflow | Hew et al. (2022), Jeon (2021,2022) |

| | |
|-------------|-----------------------|
| Replika | Lin, & Mubarak (2021) |
| Sister Fish | Liu et al. (2022) |
| CSIEC | Ebadi, & Amini (2022) |
| TPBOT | Hsu et al. (2021) |

In the reviewed studies, chatbots were used to provide learners with various pedagogical affordances to support and facilitate their learning process. In Yang et al.'s (2022) study, Ellie, a voice chatbot based on pedagogic tasks, was used by students to complete three interaction tasks such as shopping, taking orders, etc. In that study, Ellie was the conversation partner of learners. In addition, El Shazly (2021) utilized various AI-driven chatbots in her study, namely Mondly and other web chatbots such as Mike, Charles, Audrey, and Cristal, to enable both spoken and written interactions possible for learners with the chatbots, and dialog-based practices were at the center of the applications. The LINE ChatBot in Chien et al.'s (2022) study was used to encourage learners to engage in English conversation activities, and thus to improve their listening and speaking practices. In Hwang et al.'s study (2022), SmartU English chatbot was designed to help learners with speaking and pronunciation fluency and to offer individual feedback. It provided learners with the support and opportunity to have both designed (i.e. given sentence patterns) and free talks (without sentence patterns) with the chatbot. Studies by Hew et al. (2022) and Jeon (2021, 2022) did not use a chatbot with a specific name, but they stated that they created and used educational chatbots built by Google Dialogflow. In Lin and Mubarak's (2021) study, students used Replika to practice English conversations and improve their speaking skills. The chatbot Sister Fish in Liu et al.'s (2022) study accompanied and interacted with learners about the English books (made up of simple stories) that learners read. In Ebadi and Amini's study (2022), the chatbot CSIEC was employed to allow learners have a conversation with the chatbot regarding job application interviews. Finally, the task-based chatbot TPBOT was used by Hsu et al. (2021) to reduce learners' speaking anxiety and help them improve their speaking skills through dialogues/questions and answers. The TPBOT was designed with a correction system that allowed learners to correct their mispronunciation by imitating and practicing what the TPBOT said.

3.3. Results of the reviewed studies

The results of the reviewed studies (see Table 7) show that chatbots can positively contribute to the language learning process of EFL students. In addition to chatbots' positive contribution to overall language achievement (see Hwang et al., 2022), 6 out of 11 studies using pre- and post-tests demonstrated that chatbots had positive effects on achievement of specific English language skills and domains such as speaking (e.g. Chien et al., 2022; El Shazly, 2021; Hsu et al., 2021; Hwang et al., 2022; Lin, & Mubarak, 2021), listening (Chien et al., 2022), and vocabulary (Jeon, 2021). Also, 4 studies that did not use pre- or post-tests but used conversation/dialog/interaction logs and interviews/surveys found that chatbots affected speaking (e.g. Jeon, 2022; Yang et al., 2022), listening (e.g. Hew et al., 2022; Jeon, 2022), and reading (e.g. Liu et al., 2022) positively. Additionally, 6 studies (e.g. Hew et al., 2022; Jeon, 2021, 2022; Lin & Mubarak, 2021; Liu et al., 2022) that included conversation/interaction logs as data collection tools revealed that students generally had effective and beneficial conversation practices with chatbots.

In addition to the findings on the effects of chatbot use on language learning success and process, most of the studies also yielded some results on affective/psychological influences of chatbot use in the language learning process. Some studies (e.g. Ebadi & Amini, 2022; Hew et al., 2022; Hsu et al., 2021;

Hwang et al., 2022; Liu et al., 2022; Yang et al., 2022) indicated that students had positive attitudes and perceptions toward the use of chatbots for EFL learning. In addition, a few of them reported that chatbots helped students maintain their willingness to communicate (see El Shazly, 2021; Jeon, 2022). One of the common findings is that chatbots can play a useful role in motivating learners in their language learning process (see Jeon, 2022; Chien et al., 2022; Ebadi & Amini, 2022; Hsu et al., 2021; Hwang et al., 2022; Liu et al., 2022; Yang et al., 2022), and in providing them with interesting learning experiences (see Hsu et al., 2021; Hwang et al., 2022; Liu et al., 2022; Yang et al., 2022). However, in Jeon's study (2022), which explored both the opportunities and limitations of chatbots, it was underlined that although some features of the chatbots were motivating for learners, some challenges that students encountered when using chatbots decreased their motivation. The reasons behind this demotivation were highlighted as chatbots' not being able to recognize L1 expressions of low language-level EFL learners, the inability of chatbots to be a substitute for human teachers, and learners' beliefs that doing activities in collaboration with their peers is more effective than doing them with chatbots (Jeon, 2022). Similarly, in Hew et al.'s (2022) study, it was stated that after a certain period of interaction with chatbots, students were able to predict the responses of the chatbots, so their interest in using them decreased. A similar case was also reported in Liu et al.'s (2022) study, in which one student said that the responses s/he received from the chatbot were similar in general, in other words, there was not much variety in responses.

Studies by Ebadi and Amini, (2022), Hsu et al. (2021), and Jeon (2022) demonstrated in their findings that chatbots can help students reduce their speaking anxiety because they do not worry about making mistakes like they do in front of their peers, and there is no one to judge them during their speech (Jeon, 2022). However, the study by El Shazly (2021) found that students' anxiety levels were similar both before and after chatbot use. In other words, they stated that the level of anxiety did not decrease after using the chatbot. According to El Shazly, one of the main reasons for this result may be that the students could not fully get accustomed to the use of chatbots as the study intervention was short, and so their anxiety level did not change.

In addition to the findings related to the psychological aspects, feedback provided by chatbots was another focus of some research. In some studies, it was stated that one of the positive features of using chatbots is that students can get instant feedback from chatbots for the utterances they produce (see Ebadi & Amini, 2022; Jeon, 2022; Hew et al., 2022; Hwang et al., 2022). However, in Hew et al.'s (2022) study, based on the comments of some participants, the need and importance of the development and improvement of chatbots in comprehending students' responses better and providing more personalized feedback to them was emphasized. Table 7 provides a summary of the reviewed articles, including the aim, intervention and main findings.

Table 7. Summary of the reviewed articles

| Study | Research aim | Intervention (EG=Experimental Group, CG=Control Group) | Key findings |
|---------------------|---|---|--|
| Chien et al. (2022) | To compare the effects of chatbots with and without peer competition features on learners' English speaking and listening skills. | The experimental research included 73 Taiwanese EFL high school students. While the EG participated in the LINE Chatbot exercises that included peer competition, the CG used the LINE Chatbot without peer competition. During the 4 weeks of the 6-week study, students used the chatbot to | Participants' speaking and listening skills notably improved thanks to the contextual learning opportunity provided by the LINE ChatBot. There was no significant difference between the two groups in terms of speaking and |

| | | | |
|------------------------|--|---|---|
| | | perform after-class learning activities based on listening and speaking. | listening ability. Both groups' intrinsic and extrinsic motivations were higher in their post-test scores. |
| Ebadi and Amini (2022) | To examine the effects of chatbot use on students' motivation. | The study included 256 university-level EFL participants from Iran. It investigated whether there were any associations between human-likeness and the social presence of the chatbot CSIEC on EFL students' motivation. Students had a 3-week experience of interaction and learning with the chatbot. | The results demonstrated that both human-likeness and social presence were significant positive predictors of student motivation. The results also showed that chatbot use for learning had a positive role in motivating students during their learning process. |
| El Shazly (2021) | To determine the role of AI-chatbots in managing learners' foreign language anxiety and improving their speaking skills. | The quasi-experimental study was carried out with 48 university-level Egyptian EFL learners. From the 2 nd week to the 7 th week of the 8-week intervention, participants interacted with various web chatbots and Mondly in written or spoken forms. | The results showed that the use of chatbots did not reduce learners' anxiety. The results also indicated that there was a significant difference between pre- and post-speaking tests, and based on this, it was suggested that chatbots can be used as effective tools to improve EFL learners' speaking skills. |
| Hew et al. (2022) | To investigate how student goal setting and social presence in online learning can be supported by chatbots. | Two case studies were conducted. In study 1, the participants (29 EFL postgraduate students) interacted with a chatbot that assisted them with their goal setting before attending their online course. In study 2, the learning buddy chatbot created in line with the social presence framework was used to help university-level EFL learners (n=38) in their listening tasks and practices. | Students indicated that they had positive experience with the use of chatbots during their learning process. |
| Hsu et al. (2021) | To examine the effects of the chatbot on EFL learners' speaking anxiety and speaking skills. | The experimental study was comprised 48 university-level Taiwanese students. The EG (n=24) used the TPBOT chatbot to practice speaking, and the CG (n=24) did not use any chatbots. | Interacting with the chatbot significantly improved learners' speaking skills. Most of the students had positive perceptions about the use of the chatbot. TPBOT had a positive role in alleviating learners' anxiety and fear of speaking English with foreigners. |
| Hwang et al. (2022) | To investigate the effects of the smart chatbot app, Smart UEnglish, on EFL learners' success and to determine learners' perceptions about the chatbot system. | The quasi-experimental study included 43 elementary school EFL learners. The students in the EG used the smart UEnglish chatbot app, however, the CG used the same system without smart mechanisms. The entire study lasted 10 weeks. | The findings demonstrated that the EG outperformed the CG in terms of learning achievement. According to the interview results, the students were interested in using the chatbot and had a positive attitude toward its use for conversation. They indicated that the smart chatbot facilitated their speaking skills. |
| Jeon (2021) | To determine the effects of chatbot-assisted dynamic assessment (CA-DA) on learners' | The study included 53 EFL learners from a primary school in Korea. There were 3 groups: a) CA-DA group, Chatbot-Assisted Non-Dynamic Assessment (CA-NDA) group, and a | The CA-DA group was found to be more successful in vocabulary learning than the other two groups. In addition, interaction data analysis results |

| | | | |
|------------------------|---|---|---|
| | productive and receptive vocabulary knowledge. | CG. While the CA-DA group read the texts and got graduated help from the chatbots to find the meaning of the target words, in the CA-NDA group, the chatbot only provided the meaning of the target words in the texts. Chatbots were not used in the CG. | confirmed the positive effects of CA-DA on the improvement of vocabulary gain. |
| Jeon (2022) | To investigate the affordances presented by chatbots and their effects on EFL learners' motivation to learn English. | The participants of the study comprised 36 primary school students in Korea. As a part of 16-week EFL class, the participants were exposed to the use of chatbots created by Dialogflow. | The results showed that there were technological, social, and pedagogical affordances, each of which was considered as either a limitation or an opportunity by different EFL learners. These opportunities and limitations led to an increase or decrease in students' motivation. |
| Lin and Mubarak (2021) | To investigate the effects of chatbots on EFL students' speaking skills. | This quasi-experimental study included 50 university-level EFL learners in Taiwan. In a flipped English speaking class, the EG was exposed to mind map-guided AI chatbot approach (MM-AI) and the CG was exposed to the conventional AI chatbot (C-AI) approach. AI-driven chatbots were used to improve learners' speaking proficiency. During the 3 weeks of the 5-week study, learners used chatbots to practice speaking. | The findings showed that the learners in the MM-AI group improved their speaking skills better than the learners in the C-AI group. When the chatbot interactive behaviors of the participants were analyzed, it was observed that the EG outperformed the CG in terms of interactive behavior frequency. |
| Liu et al. (2022) | To investigate whether the use of chatbots affects learners' interest and involvement in reading. | This study investigated the interaction of elementary school young EFL learners (n=68) with the chatbot about the books they read. The role of the chatbot was to facilitate learners' reading by interacting with them about the target book. The EG had an interaction with the chatbot named Sister Fish about the book they had just read. The students in the CG did not talk to the chatbot about the book they read. | The findings showed that interacting with the chatbots about the books stimulated and sustained students' interests in reading, and students had a positive reading experience. The perception regarding social connection with the chatbot was high. The results indicated that chatbots can be integrated into the language learning process as book talk companions. |
| Yang et al. (2022) | To investigate how a voice chatbot can be used as a conversational partner to help learners improve their conversation in speaking classes. | The study included 314 EFL learners (n=177 from elementary school; 137 from high school) in Korea. Participants had conversations with the chatbot to complete the three given conversation tasks (e.g. taking an order at a restaurant). | The results indicated that the chatbot motivated learners to take part in conversation and complete the given tasks. In other words, the integration of the chatbot helped learners improve their conversation. Most learners had positive impressions about the use of the chatbot in their learning process. |

4. Discussion

This study reviewed recent empirical studies on the pedagogical use of chatbots in EFL learning settings. After reviewing the articles, in terms of research features, it was found that the number of studies

conducted in Taiwan (n=4) was higher than in other countries. This result is consistent with that of Hwang and Chang (2021), who reported that Taiwan (n=4) was the place where chatbot-related publications in education were most common after the United States (n=7). This finding shows the necessity of increasing the number of studies in various countries to obtain more comprehensive information about chatbot use in the EFL teaching and learning domain. As found in previous review studies (e.g. Zhai & Wibowos, 2022), it was identified that different types of chatbots with various features were used for pedagogical purposes in the reviewed papers. In addition, the fact that most of the studies included a mixed-method design made it possible to evaluate the study findings from different aspects. With regard to the education context, the number of studies focusing only on higher education was higher. A similar finding was also reported by Lee and Hwang (2022). To have a clearer picture of the impacts of chatbot applications, more studies at various levels of education are needed.

Considering the results of the examined articles, it can be deduced that chatbots can have a positive effect on EFL learning achievement. This finding agrees with previous review and meta-analysis studies such as Lee and Hwang (2022), Deng and Yu (2023), and Haristiani (2019), which also underlined that chatbots had positive effects on learning success. Additionally, different chatbots used in the studies offered various pedagogical affordances to the students, as in Huang et al.'s (2022) study. Although chatbots were found to contribute positively to the language learning process in general, some studies also shared the limitations and problems related to their use and application. Feedback is one example of this situation. On the one hand, as stated before by some studies, chatbots provide effective and fast feedback. This corroborates previous review studies (e.g., Fryer & Carpenter 2006), which also support that chatbots are good feedback providers. On the other hand, in some studies in the current review, it was underlined that some users reported that chatbots were insufficient to understand their utterances very well and to give personal feedback to them (e.g. Hew et al., 2022). This issue was also underlined in a previous review study by Jiang (2022). Jiang's study highlighted that chatbots may have difficulty detecting and dealing with some mistakes in language such as grammar, and spelling. Also, the problem that chatbots cannot recognize some utterances of learners supports the findings of the Huang et al.'s (2021) review study. In addition, the findings related to the limited and predictable responses of chatbots in some studies (e.g. Hew et al., 2022; Liu et al., 2022) is in line with the findings of the review by Haristiani (2019) and Huang et al. (2021) in terms of the technological limitations of chatbots. Therefore, on the basis these studies, it is suggested that chatbots should be developed in a way that allows them to better understand what users say and write, and provide more effective feedback and various responses to them. Also, this review study examining studies with different research designs found that experimental studies mostly focused on the effects of chatbots on learners' speaking skills. However, in Huang et al.'s (2022) review study, the reviewed experimental studies (n=8) were mostly on listening, writing, and grammar. Based on this, it can be said that there is a need to increase the number of experimental studies on the use and effects of chatbots for different language skills, such as reading and writing.

There are also some points that need to be discussed in terms of affective or psychological dimensions in the use of chatbots in the context of EFL. For instance, it was stated in the findings of many reviewed studies that the use of chatbots in language learning had a positive influence on students' motivation. Contrary to this finding of the present study, in their meta-analysis study, Deng and Yu (2023) reported that chatbots did not have notable effects on learner motivation. These contradictory results necessitate more detailed research on the effect of chatbots on student motivation. Besides, it is important to improve the features of chatbots in line with the interests and needs of the students to maintain their motivation. Another topic of discussion is the effects of chatbots on EFL learners' anxiety when

communicating with chatbots. While three reviewed studies underlined that chatbots can alleviate learners' anxiety, one study underlined that chatbots had no significant effect in reducing anxiety (see results of the reviewed studies section). Related to this issue, Zhai and Wibowo's (2022) systematic review revealed that chatbots that are designed and equipped with empathetic dimensions were found to be effective in eliminating foreign language anxiety of second language learners. Based on this, integrating chatbots with affective or empathetic functions in studies to reduce or eliminate foreign language learners' anxiety will enable us to obtain clearer results.

5. Limitations

There are some limitations of this study. First, as this study is limited to a small number of studies, the generalizability of the results obtained cannot be guaranteed. Second, the study only included SSCI-indexed empirical articles from the WoS database, and studies that only focused on EFL/ESL learning contexts. This also makes the findings less generalizable to other L2 learning contexts.

6. Conclusion and suggestions for future research

In this systematic review study, eleven studies on the use of chatbots in the EFL learning contexts were analyzed. The main contribution of this study to the literature is that it presents the recent empirical research applications related to the integration of chatbots in EFL learning process. It also contributes to the field and sheds light on the future research by reporting both the main findings of the studies and the types of chatbots used, and the characteristics of the research of each paper. In addition to showing how different types of chatbots were applied in research, this review also provided the opportunity to see the advantages and some of the challenges of integrating chatbots into the EFL learning settings to support learners.

The reviewed studies suggest that chatbots offer considerable opportunities for EFL learners to practice the target language. In all studies, it has been underlined that chatbots make a positive contribution to the EFL learning process. In addition to the positive outcomes, only a few studies mentioned the negative experiences of learners related to the use of chatbots. Since the active use of chatbots in the field of foreign language education is relatively new, the observed and experienced deficiencies need to be improved. In the light of developing technology, experts (e.g. chatbot developers), practitioners, and educators can work together to offer possible solutions and can collaborate to overcome the aforementioned limitations.

In line with the findings, this study also has some suggestions for further research. In this review, it was identified that chatbot studies in the field of EFL education mostly focused on speaking or conversation skills; therefore, it will be an interesting subject for future studies to explore the impact of chatbot-embedded education on the development of different language skills and domains such as writing, reading, and vocabulary. Additionally, future studies should focus on how the limitations of chatbots mentioned in the findings can be improved for effective language education. Moreover, future research can also explore the effectiveness of chatbot use for different proficiency levels (e.g. beginner, intermediate, advanced, etc.) of EFL learners to better understand the influences of chatbot use. Further studies could assess the long-term effects of chatbot use in the EFL context, which will allow us to see the advantages and disadvantages of using chatbots in EFL contexts more comprehensively. Finally, as this study focused on three major factors (i.e. research features, used chatbots, and findings) in the

articles, it is recommended that future review studies examine chatbot related articles from different perspectives.

References

*Reviewed articles

- Bhise, A., Munshi, A., Rodrigues, A., & Sawant, V. (2023). Overview of AI in education. In P.P. Churi, S. Joshi, M. Elhoseny, & A. Omrane (Eds.), *Artificial intelligence in higher education: A practical approach* (1st ed.) (pp. 31-62). CRC Press.
- *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. <https://doi.org/10.3389/fpsyg.2022.785752>
- Chu, S. T., Hwang, G. J., & Tu, Y. F. (2022). Artificial intelligence-based robots in education: A systematic review of selected SSCI publications. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2022.100091>
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), 2940.
- *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*, 1-19. <https://doi.org/10.1080/10494820.2022.2096638>
- *El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8-14.
- Han, D. E. (2020). The effects of voice-based AI chatbots on Korean EFL middle school students' speaking competence and affective domains. *Asia-Pacific Journal of Convergent Research Interchange*, 6, 71-80.
- Han, J. W., Park, J., & Lee, H. N. (2022). Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: A quasi-experimental study. *BMC Medical Education*, 22, 830.
- Haristiani, N. (2019, November). Artificial Intelligence (AI) chatbot as language learning medium: An inquiry. In *Journal of Physics: Conference Series* (Vol. 1387, No. 1, p. 012020). IOP Publishing.
- *Hew, K. F., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in Higher Education*. <https://doi.org/10.1007/s12528-022-09338->
- *Hsu, M. H., Chen, P. S., & Yu, C. S. (2021). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 1-12. <https://doi.org/10.1080/10494820.2021.1960864>
- Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 1-14.
- *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*, 1-27.
- *Jeon, J. (2021): Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2021.1987272>

- *Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 1-26.
<https://doi.org/10.1080/09588221.2021.2021241>
- Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, 13.
<https://doi.org/10.3389/fpsyg.2022.1049401>
- Lee, J.-Y., & Hwang, Y. (2022). A meta-analysis of the effects of using AI chatbot in Korean EFL education. *Studies in English Language & Literature*, 48, 213–243.
- *Lin, C.-J., & Mubarak, 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, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, 189, 104576.
- Pereira, J., Fernandez-Raga, M., Osuna-Acedo, S., Roura-Redondo, M., Almazan-Lopez, O., & Buldon-Olalla, A. (2019). Promoting learners' voice productions using chatbots as a tool for improving the learning process in a MOOC. *Technology Knowledge and Learning*, 24(4), 545–565.
<https://doi.org/10.1007/s10758-019-09414-9>
- Perez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549–1565.
- Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, 151, 103862.
- Xia, L., & Zhong, B. (2018). A systematic review on teaching and learning robotics content knowledge in K-12. *Computers & Education*, 127, 267–282. <https://doi.org/10.1016/j.compedu.2018.09.007>
- *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>
- Yasmin, H. & Mazhar, R. (2023). AI in education: A few decades from now. In P.P. Churi, S. Joshi, M. Elhoseny, & A. Omrane (Eds.), *Artificial intelligence in higher education: A practical approach* (1st ed.) (pp. 1-30). CRC Press.
- Zhai, C., & Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: The case of second language acquisition. *Heliyon*.
<https://doi.org/10.1016/j.heliyon.2022.e12056>