

The Educational Effects in Language Learning of Chatbots

Shisen Yue

Department of Linguistics, Shanghai Jiao Tong University

Abstract

Practicing conversational skills with chatbots has become an innovative method in language learning. The AI-powered communication partners have demonstrated multiple advantages compared to human partners, such as being less time-consuming and less anxiety-induced. Moreover, students' interest in the innovative learning tool can be converted into a motivation in language learning. However, there is also space for chatbots to improve for better learning experiences and effects. To gather the positive and negative effects of involving chatbots in language learning, we review a list of empirical studies on this topic. Also, we analyze the studies with a unified framework and examine their conclusions.

The Educational Effects in Language Learning of Chatbots

Introduction

As the product of Artificial Intelligence (AI) development, conversational agents, or chatbots, have taken up an assistantship in language learning. The AI-powered dialogue systems provide students with access to the native-level mastery of the language at every moment. The design of chatbots can be tailored to a specific pedagogical goal, such as grammar improvement or pronunciation correction. Students also demonstrate a decreased speaking anxiety and enhanced motivation in L2 learning (Jeon, 2021; Tai & Chen, 2023)

Comprehensive evaluations of the positive effects of the involvement of chatbots in language learning have been conducted (e.g. Kohnke, Moorhouse, and Zou 2023; Shadiev and Liu 2023; Zhang, Zou, and Cheng 2023). The conclusions are drawn from experiments that test the learning effects of human participants after training with chatbots with different characteristics. Some chatbots are specifically created for the experiment (Jeon, 2021), while others are commercial products for general use, such as Google Assistant or Alexa (Dizon, 2020). Moreover, some chatbots have a virtual, human-like appearance (Divekar et al., 2022), while others operate without a visual configuration (Xu, Wang, Collins, Lee, & Warschauer, 2021). The diversification in chatbots is credited to the rapid development of language technology.

The variation in chatbots necessitates a unified framework to describe their characteristics to enhance the generalizability of the chatbots' educational values drawn from empirical studies. According to the perspective from which they describe the chatbots, the existing frameworks can be classified as design-centered (Bibauw, François, & Desmet, 2019; Jeon, Lee, & Choe, 2023; Kim, Cha, & Kim, 2019) or pedagogical (Huang, Hew, & Fryer, 2022; Ji, Han, & Ko, 2023). The former approach focuses primarily on identifying different types and features of chatbots for language learning, and the latter mainly centers on exploring the way chatbots are utilized.

In the current paper, we review 12 selected studies evaluating the application of chatbots in language learning, in order to (1) summarize the advantages and disadvantages of applying the chatbots to language learning and (2) examine the conclusions about the chatbots' educational effects from a design perspective.

Method

A comprehensive review of studies that evaluate the application of different chatbots in language learning is conducted. We use Google Scholar as the primary engine to collect and screen the studies. To gather all related work, we list all alternative expressions, including the hypernyms and hyponyms of “chatbots” and “language learning” and form different combinations between them as keywords for search. Specifically, “chatbots” can be replaced with “conversational agents”, “conversational systems”, “intelligent dialogue systems” and “interactive AI”, and “language learning” alternates with “EFL”, “language acquisition”, and “foreign language”. By virtue of the searching method of Google Scholar, we gathered a list of papers whose title, abstract, or keywords involve any of the strings in the search keywords.

The studies are analyzed with the GEM framework (Jeon et al., 2023). It includes three descriptive features of the chatbots, namely goal orientation, embodiment, and multimodality. The framework was established by first selecting from the general features shared by different types of conversational technology, such as embodiment (eg., Li 2015; Xu et al. 2021), multimodality (eg., Morton and Jack 2010; Tai 2022), pedagogical roles (eg., Ji et al. 2023; Kuhail, Alturki, Alramlawi, and Alhejori 2023), physicality (eg., Li 2015) and form-meaning constraints (eg., Bibauw et al. 2019). Then the authors append the specific features discussed in each of the studies they reviewed. They exclude the features related to pedagogical applications, leaving only three features that describe the design of chatbots. According to Jeon et al. (2023), goal orientation suggests whether a chatbot is designed to satisfy the needs of a specific group of language learners, and embodiment indicates whether a chatbot has a virtual body or similar demonstration of being a chatting entity instead of a computer program, and multimodality reflects if multiple communication channels, such as text, voice, and videos provided to enrich the experience of users.

Instead of appearing as features, the pedagogical applications are regarded as the derivative effects of the three features within the GEM framework. The bridge between the design of chatbots and their real-world applications is established upon the accordancy theory proposed by Norman (1988), which emphasizes the design of objects as a crucial determinant of the affordances, the quality of the learning experience under the current context. By reviewing the studies that evaluate chatbots’ application in language learning, Jeon et al. (2023) summarizes the pedagogical applications that can be derived from each of the three

features in the GEM framework. Specifically, the goal-orientated chatbots are related to (1) situational tasks, (2) instructional scaffolding, (3) varied task difficulty, and (4) flexible customization, the chatbots that have embodiment are associated with (1) immersive environment and (2) personal bond formation, and those with multimodality can help facilitate (1) self-monitoring and self-correction, (2) comprehension support, and (3) exploration of information. While the chatbots that carry certain features can create the corresponding educational effects, the actual effects vary by their specific designs and the experimental setting.

Results

We review 12 studies about the application of intelligent chatbots in language learning (See Appendix for detailed information). These studies reveal the strengths and drawbacks of improving the language proficiency level with the conversational agents, examining the possibility of integrating them into mainstream pedagogy in language teaching. The studies also discuss the difference between interacting with a chatbot and a human conversation partner and how this difference might influence the student's attitudes and language improvements.

The improvement in L2 learners' language skills after practicing with chatbots is represented by the increased vocabulary (Divekar et al., 2022), enhanced conversational skills (Dizon, 2020; H.-L. Hsu, Chen, & Todd, 2021), less grammatical errors and pronunciation duration (Hassani, Nahvi, & Ahmadi, 2016). While students demonstrate enhanced skills in speaking the target language, their listening skills do not improve significantly (Dizon, 2020; M.-H. Hsu, Chen, & Yu, 2023). These results suggest that the conversational agents are useful tools targeted at the speaking ability of L2 learners.

Compared to a native speaker, a chatbot induces less anxiety from the L2 learners during the conversation. Especially, low- and intermediate-level students feel more comfortable expressing their ideas in front of a non-human agent (Divekar et al., 2022). Moreover, the chatbots add fun to the learning process by virtue of the involvement of multimodal information and the humorous responses (Ayedoun, Hayashi, & Seta, 2019; Dizon & Tang, 2020; Gonulal, 2023). The chatbots also demonstrate the ability to assess L2 learners' English proficiency level comparable to human interviewers (Forsyth et al., 2019).

There are also drawbacks to practicing with chatbots. First, communication breakdown happens when low- and intermediate-level learners converse with them (Chen, Yang, & Lai, 2023). Since chatbots lack a pre-designed program to solve the breakdown, the students mostly repeat the sentences to repair the conversation. This could lead students to give up the training (Dizon & Tang, 2020). Furthermore, students' interest in this state-of-the-art teaching approach easily fade away when they get accustomed to it (Fryer, Ainley, Thompson, Gibson, & Sherlock, 2017). An out-of-class setting where professional instruction from the teacher is unavailable also dampens students' passion for practicing with chatbots (Dizon & Tang, 2020). However, Fryer, Nakao, and Thompson (2019) suggests that the students' interest in chatbots generally experience a 20-week interval to rebound back. They also indicate that this interest is linked to the individuals' willingness to have conversations with human partners.

Discussion

The reviewed studies though established their conclusions upon adequate empirical evidence, might be based on the wrong choice of chatbots. Their conclusions might be therefore unable to generalize over all chatbots in their compatibility to L2 learning. Chatbots vary in the goal they are designed for and therefore their affordances. Given that interlocutors with different characteristics provide different interaction experiences even on the same topic, the choice of chatbots should be aligned with the aim and the scope of the research.

Through the lens of a generalizable framework in describing the characteristics of chatbots (Jeon et al., 2023), we can identify more easily whether the chatbot is suitable for the study. For instance, utilizing Alexa, a commercial chatbot that does not have embodiment and multimodal information apart from voice, Dizon and Tang (2020) proposes that students' interest in chatbots decreases after class. While this conclusion is reasonable for the chatbot Alexa, it might not apply to other chatbots, especially those with embodiment with vivid animations and information provided in other channels. Divekar et al. (2022) find very positive feedback from students in practicing with a customized chatbot which is embodied by a 3D animation of a market shopkeeper and outputs information in text, voice, image, and user-gestures. However, the abundance in the form of communication could play a role even more crucial than the responses from the chatbot in arousing students' interests. The estimate of the interval for students to recover their interest (Fryer et al., 2017) could also be

chatbot-specific.

While the GEM framework systematically describes the design and the pedagogical applications of the chatbots, the technical details in training the language model underlying a chatbot are not taken into consideration. Similar to the design of a chatbot, the approach and data used for training a language model decide the scenario it's suitable for. For instance, a chatbot built on language models trained with encyclopedic knowledge crawled from the internet might have less support in developing L2 learners' pragmatic knowledge of the target language.

Conclusions

In this paper, we review 12 selected studies that evaluate the educational effects of involving intelligent chatbots in language learning within a framework proposed by Jeon et al. (2023). The results indicate that chatbots are useful tools in assisting language learning due to their non-human identity and multiple channels of information. The studies also revealed the disadvantages of applying chatbots in language teaching, which include the inability to tackle the conversation breakdown and the non-persistent interests from students. Moreover, we reveal the necessity of a systematic framework that describes both the design of chatbots and the underlying language models in chatbot research.

References

- Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, *29*, 29–57.
- 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.
- Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college efl learners' perceptions toward the use of google assistant for foreign language learning. *Interactive Learning Environments*, *31*(3), 1335–1350.
- Divekar, R. R., Drozdal*, J., Chabot*, S., Zhou, Y., Su, H., Chen, Y., . . . Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning*, *35*(9), 2332–2360.
- Dizon, G. (2020). Evaluating intelligent personal assistants for l2 listening and speaking development.
- Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of alexa. *JALT CALL Journal*, *16*(2), 107–120.
- 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.
- 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.
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in human Behavior*, *93*, 279–289.
- Gonulal, T. (2023). Investigating efl learners' humorous interactions with an intelligent personal assistant. *Interactive Learning Environments*, *31*(7), 4521–4534.
- 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.

- 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*, 1–14.
- Hsu, M.-H., Chen, P.-S., & Yu, C.-S. (2023). Proposing a task-oriented chatbot system for efl learners speaking practice. *Interactive Learning Environments*, 31(7), 4297–4308.
- 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.
- Jeon, J. (2021). Exploring ai chatbot affordances in the efl classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 1–26.
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond chatgpt: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0360131523001756> doi: <https://doi.org/10.1016/j.compedu.2023.104898>
- 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.
- Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future english learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, 22(3).
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Chatgpt for language teaching and learning. *RELC Journal*, 00336882231162868.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973–1018.
- Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77, 23–37.
- Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: A cross-cultural evaluation. *Computer Assisted Language Learning*, 23(4), 295–319.
- Norman, D. A. (1988). *The psychology of everyday things*. Basic books.

- Shadiev, R., & Liu, J. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, 1–15.
- Tai, T.-Y. (2022). Effects of intelligent personal assistants on efl learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 1–30.
- Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of google assistant on adolescent efl learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502.
- 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.
- Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*, 1–29.

Appendix

Study and main findings	GEM feature information			Pedagogical application
	G	E	M	
Ayedoun et al. (2019) found learners were more pleased and willing to communicate with the chatbot that employed a combination of communication strategies and affective backchannels than those using only one strategy of the two.	✓ Customized chatbot for L2 learning	✓ 3D animation (restaurant waiter avatar)	✓ text	Situational task, instructional scaffolding
Chen et al. (2023) found L2 learners considered the chatbot an easy tool to improve their language skills. Communication breakdowns happened on learners at low and intermediate proficiency levels, and repetition is a common way to repair.	✗ Commercial chatbot for general use (Google Assistant)	✗	✓ text, image	Self-monitoring and self-correction
Divekar et al. (2022) found that chatbots were useful tools in enhancing L2 students' vocabulary, comprehension, and conversation skills. The chatbots created a comfortable and less-anxiety-inducing space for students to practice in the target language.	✓ Customized chatbot for L2 learning	✓ 3D animation in a virtual space (market shop-keeper avatar)	✓ text, image, user gesture	
Dizon & Tang (2020) found that students showed interests to chatbots for the fun, easy-to-use and effective way to learn English. However, their enthusiasm faded in an out-of-class setting, and they easily gave up when communication breakdown occurred.	✗ Commercial chatbot for general use (Alexa)	✗	✗	Not found
Dizon (2020) found the chatbots is capable in improving students' L2 speaking proficiency but not L2 listening proficiency.	✗ Commercial chatbot for general use (Alexa)	✗	✓ text, image	Not found
Forsyth et al. (2019) found the chatbots have the ability to assess L2 learners' English abilities comparable to humans interviewers	✓ Customized chatbot for L2 learning	✓ 3D animation (teacher and student avatar in a virtual space)	✗	Situational tasks, instructional scaffolding, personal bond formation
Fryer et al. (2017) found L2 learners' enthusiasm could easily fades away, unlike the sustained interest obtained from communicating with human partners.	✗ Commercial chatbot for general use (Cleverbot)	✗	✓ text	Not found
Fryer et al. (2019) found L2 learners' decreased interest in chatbots recovered after a 20-weeks interval. Additionally, learners' interest in conversing with chatbot was linked to interest in communicating with human partners and the improvement is correlated with task interest.	✗ Commercial chatbot for general use (Cleverbot)	✗	✓ text	Not found
Gonulal (2023) found L2 learners enjoyed the humorous responses from chatbots and thus perceived them as useful tools.	✗ Commercial chatbot for general use (Google Assistant)	✗	✓ text, image	Not found
Hassani et al. (2016) found L2 learners made more proper responses and less grammatical errors and had less pronunciation duration through practicing with chatbots.	✓ Customized chatbot for L2 learning	3D animation in a virtual space (sales agent avatar)	✓ user gesture	Situational task, varied task difficulty
M.-H. Hsu et al. (2023) found that the training with chatbot doesn't enhance learners' listening proficiency.	✗ Commercial chatbot for general use (Alexa)	✗	✗ text, image	Comprehension support
H.-L. Hsu et al. (2021) found that using the chatbots improved students' speaking skills.	✓ Customized chatbot for L2 learning	✗	✓ text	Instructional scaffolding

Table 1

The detailed information of reviewed studies. Supplementary text in the column M indicates the multimodal approach apart from voice.