An investigation of the efficacy of a ChatGPT-powered chatbot

Michael Wilkins[a], Mark Donnellan[b] and Kane Linton[c]

[a]The Language Center, Kwansei Gakuin University, iD, michaelwilkins@kwansei.ac.jp; [b]The Faculty of Informatics, Kindai University, iD, donnellan@kindai.ac.jp and [c]School of Engineering, Kwansei Gakuin University, iD, kane.linton.1985@gmail.com

Abstract

One of the newest tools for language learning is AI-powered chatbots which allow students to engage in interactive conversations with a chatbot. English Central is a digital language learning platform that allows students to watch videos, study vocabulary, and produce spoken language through short interactive videos. In addition to these three core components, English Central has recently added a chatbot powered by ChatGPT. This paper shares the results of a pilot study that analyzed data from Japanese university students who engaged in interactions with this chatbot. The authors introduce quantitative data showing the quantity and quality of the interactions and discuss the strategies they implemented to encourage longer interactions. A student feedback survey was conducted and the results of this are also shared. The results suggested that although many students reported benefits to using a chatbot, the strategies implemented did not lead to significantly longer interactions, and that many students felt frustrated when the chatbot did not understand their utterances.

Keywords: *Chatbot, ChatGPT, English Central, AI*

1. Introduction

English Central (EC) is a digital language learning platform that allows students to do three core activities: watch videos; learn words; and speak lines from the videos. Language learners interact with these short videos getting invaluable listening and pronunciation practice and opportunities to learn vocabulary in context. In addition to these three core components, EC has recently integrated a chatbot called Mimi, powered by ChatGPT technology. For the purposes of this study, the researchers are specifically interested in students' interactions with Mimi. This paper reports on data from those interactions, shares student feedback on Mimi, and offers suggestions on how teachers can facilitate more meaningful interactions between chatbots and students.

Although the use of chatbots for language learning is considered by many to be a completely new field, there is already a well-established and rapidly growing body of supporting literature, including numerous review and meta-analysis articles (Fryer et al., 2020; Huang et al., 2022; Klímová & Ibna, 2023; Wollny et al., 2021; Zhang et al., 2023). Some researchers, such as Fryer & Carpenter (2006), have been working on using chatbots for language learning for almost 20 years. Key areas of inquiry into chatbot use have been: the affordances of chatbots (Wollny et al., 2021); instructional design and application (Fryer et al., 2020); learner objectives, experiences, and

challenges (Okonkwo & Ade-Ibijola, 2021); evaluation (Pérez et al., 2020); and technology design (Chen et al., 2023).

Chatbots are well-established as conversational agents to answer customer questions for online businesses and millions of people have interacted with digital assistants, such as Apple's Siri, showing a general acceptance of interacting with non-human entities. However, for language learning the results had not been as impressive (Fryer et al., 2020). With the recent advancements in AI, such as neural network language models, natural language processing, and the use of larger language model datasets, chatbots such as Open AI's Chat GPT 3.5 can interact with users in a more genuine and responsive way (Hong, 2023). However, to date, only a few studies have analysed the characteristics of the actual language produced by students while conversing with chatbots (Kim et al., 2021) and this study will attempt to address this gap.

ChatGPT and other AI tools are recognised to be massive disruptors of the language learning industry, but their effects are only just beginning to be felt (Hong, 2023). There are many articles assessing the pedagogical effects of chatbots, such as Klímová and Ibna (2023), but most are meta-analyses of the overall effect or evaluate a single use case rather than comparing different ways to integrate an existing chatbot into a classroom setting to achieve maximum student learning. There were no studies that looked specifically into the length and characteristics of actual student responses to chatbots in language learning. This study will explore optimal chatbot use in classroom settings.

A major concern for educators introducing new technology into the classroom is how students will react. If students do not perceive chatbots to be useful, entertaining, or both they will not use them except under compulsion from the teacher, which is not conducive to student learning. Huang (2022) asserts that chatbots can lower learners' affective filter and improve motivation as well as engagement in language learning. This study aims to expand on the findings that chatbots boost student motivation and comfort. The research questions were:

RQ 1: How long were typical student interactions with Mimi in terms of words per session?

RQ 2: What effect did the pedagogical interventions have on the interactions?

RQ 3: What are students' perceptions of the usefulness of chatbots?

2. Method

2.1. Context and participants

The participants were 26 students at a university in the Kansai region of Japan who used EC as a component of their bi-weekly four-skills English class during the spring 2023 semester. Each class in the 14-week semester was 100 minutes. The students used the three core components of EC, but also the EC chatbot Mimi was introduced.

2.2. Teaching interventions

The instructor took different pedagogical stances at different stages of the semester. For weeks 1 to 4 of the semester, students used EC without explicit pedagogical support from the teacher. They were given instructions, technical support, goals, and some encouragement to use Mimi, but no explicit 'teaching' was done. Marks were awarded for the core EC tasks of watching videos, vocabulary study, and pronunciation practice, but none for chatbot interactions. Students worked exclusively outside of class with just a short mention of the goals at the start and a reminder at the end of each class. In weeks 5-8, the teacher highlighted and praised the top-performing students and offered increased encouragement. A few marks were introduced for completing chats in addition to the regular marks for EC with work still being done exclusively outside of class time. In weeks 9-12, the process was like weeks 5-8 but the number of marks for completing chats was increased and extra marks were given for the quality of the chats. The teacher listened to one chat from each student and rated it from 1 to 5 points. The EC activity was still exclusively considered to be a homework task. In week 13 the teacher took a more direct role (intervention 1), and students were given significant time in class to watch EC videos and have conversations with

Mimi. In week 14 (intervention 2) the teacher gave a demonstration of a model chatbot interaction. It consisted of a five-turn chat with Mimi with each of the five 'student' turns being several sentences long and asked the students to follow that model. As in week 13, students were given significant time in class to do their EC video-watching and chat tasks.

2.3. Chatbot Data

The chatbot data for the whole semester was downloaded. This data included all utterances by both Mimi and the students, the number of turns in each interaction, and the average number of words per interaction. Mimi allows students to give both spoken and written responses.

2.4. Feedback Survey

Following the completion of intervention 2, a survey was given to gather student feedback on students' perceptions of EC, there were 13 6-point Likert scale items pertaining to chatbot usage, and qualitative data was collected through two open-ended questions to gather positive and negative student feedback on their interactions with Mimi. The student responses were coded thematically using NVivo.

3. Results

3.1. Chatbot Data

The data below shows a summary of the pertinent data from students' interactions with Mimi during three periods: pre-intervention 1; intervention 1 to intervention 2; and intervention 2 to the end of the semester.

Table 1. Summary of student interactions with the chatbot.

	Average no. of turns per session	Average no. of words per learner turn	Average no. of words per conversation	Percentage of spoken responses
Pre-intervention	4.67	10.85	50.66	57.9
Intervention 1 to intervention 2	4.94	10.56	52.16	63.9
Intervention 2 to end of semester	4.84	10.47	50.67	60.2

3.2. Survey results

The table below shows the responses to the 13 items on the survey that related to Mimi:

Table 2. Survey results.

	M	SD
Interacting with the English Central Chatbot helped me learn English.	3	1.5
English Central and its chatbot help English learning because they are available 24 hours a day.	2.3	1.4
English Central and its chatbot help English learning because they are available anywhere.	2	1.1
Speaking to the chatbot in English is less stressful than speaking to a human in English.	3.5	1.8

The chatbot interacted with me in a human-like way.	3.2	1.2
The chatbot seemed knowledgeable and did not say obviously false things.	2.6	1.2
The chatbot gave me good feedback on my English.	3.2	1.4
I could understand the chatbot easily.	2.5	1.3
I think speaking with the chatbot helped me speak English more fluently.	2.8	1.2
The teacher's request for students to use the chatbot motivated me to use the chatbot.	2.5	1.3
The only reason I used the chatbot was because I can get some points for this class.	2.7	1.4
The teacher using class time to explain the chatbot motivated me to use the chatbot more.	2.7	1.2
I will seek out opportunities to use chatbots to learn languages in the future.	3.2	1.2

Note. N=26 (1 completely agree – 6 completely disagree)

3.3. Open-ended responses

The students responded to two open-ended questions at the end of the survey about the positive and negative aspects of using the chatbot. The results of the thematic coding can be seen below in Table 3. Overall, there were 24 positive comments and 25 negative comments.

Table 3. Thematic coding of open-ended responses.

	Number of Comments
Positive	
Authentic Communication	1
Ease of Use	2
Generally Positive	12
Improved Skills (other than speaking)	2
Improved Speaking Skill	5
Reduced Anxiety	2
Negative	
Generally Negative	6
Inflexibility	1
Lack of Feedback	3
Miscomprehension	15

4. Discussion

Chatbots and AI language learning tools are relatively new, and teachers are examining how they can be best used for learning. This study set out to measure how much language students produce when interacting with chatbots, if teacher intervention increases this amount, and how students perceive this learning activity and tool.

The results show students took four to five turns of about ten words each for conversations of about 50 words in total per chatbot interaction. It is difficult to quantify this amount of student language production as sufficient or not. Since there are no studies to establish a previous baseline length or volume of student responses to chatbot prompts in language learning, an analysis of the results must be self-referential.

The results show that, regarding RQ1 and RQ2, the average number of turns increased and the average number of words per turn decreased slightly following the interventions. Also, following the interventions, the percentage of spoken responses increased. Given that the number of turns increased, the slight decrease in the word count is not surprising. The teacher interventions did not lead to significant changes in the number of words spoken by students per conversation with the chatbot, only slight improvements. There was no data in the survey comments to explain this, but it is possible that explicit teaching earlier in the semester would lead to more student output.

Regarding RQ3, student attitudes to chatbots were mixed. The positive aspect of Mimi always being available regardless of time and place was very popular and matches the consensus of previous research (Fryer & Carpenter, 2006; Kim et al., 2021) that one of chatbots' main affordances is their convenience regarding time and place. A second often-mentioned benefit was a lowering of the affective filter when interacting with chatbots (Kim et al., 2021; Klímová & Ibna, 2023). However, this is only partially reflected in the results. While some students found talking to Mimi less stressful with one student saying, "I am not nervous when I talked with him", others at the very least found the experience frustrating as is illustrated by the student comments regarding miscomprehension below. Chen et al. (2023) contend that chatbots may not understand student input or may not behave in a human-like way. The survey data supported this, with 15 students mentioning miscomprehension as being a major frustration when using Mimi. One student commented, "I hope their speech-to-text technology will be improved", with three others saying that Mimi "didn't catch my English correctly". While speech recognition technology has progressed and one student did comment that, "The chatbot can listen [sic] my English correctly", this student was in the minority with many of the responses suggesting that improved speech recognition is essential if Mimi and other chatbots are to become prominent tools in language learning.

5. Conclusions

This study showed that students were able to interact with Mimi and have a generally positive experience but with some complaints about miscomprehension and inflexibility in Mimi's speech recognition. This will most likely improve soon as the technology advances. Both positive and negative student comments are in line with the previous literature, so there seems to be a consensus forming regarding the issues surrounding chatbots in language learning. Surprisingly, teacher interventions did not greatly influence the quantity of student responses. In the next iteration of this research, the researchers hope to increase the number participants and discover what pedagogical interventions do increase student engagement with chatbots.

References

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers*, *25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, *10*(3), 8-14.

Fryer, L., Coniam D., & Carpenter, R. (2020) Bots for Language Learning Now: Current and Future Directions. *Language Learning & Technology*. 24(2) pp 8-22.

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1).

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

Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. 영어학, *21*, 712-734.

Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology*, *14*, 1131506. https://doi.org/10.3389/fpsyg.2023.1131506

Okonkwo, C. W., & Ade-Ibijola, A. (2020). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033.

Pérez, J. Q., Daradoumis, T., & Marquès Puig, J. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. https://doi.org/10.1002/cae.22326

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence*, *4*, 654924. https://doi.org/10.3389/frai.2021.654924

Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 1-21. DOI: https://doi.org/10.1007/s10639-023-11805-6