**Table 1. Number of papers reviewed and analyzed by journal name and year of publication**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Venue** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **2024** | **Total:** |
| International Journal of Artificial Intelligence in Education | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** | **-** | **-** | **-** | **-** | **-** | **2** |
| IEEE Transactions on Learning Technologies | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| Interactive Learning Environments | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **2** | **3** | **-** | **7** |
| TESOL Journal | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** |
| Language Learning & Technology | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** | **-** | **-** | **2** |
| JALT CALL Journal | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| Expert Systems | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| Computers in Human Behavior | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **2** |
| Learning and Individual Differences | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| Computer Assisted Language Learning | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **2** | **2** | **1** | **-** | **5** | **-** | **1** | **11** |
| Applied Sciences | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** | **-** | **2** |
| **Total:** | **1** | **-** | **-** | **1** | **-** | **-** | **-** | **2** | **3** | **3** | **6** | **1** | **9** | **4** | **1** | **31** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Venue** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **Total:** |
| Novitas-ROYAL | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** |
| Education and Information Technologies | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** |
| Education Sciences | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| Educational Technology & Society | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **1** |
| Multimedia-Assisted Language Learning | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** |
| ReCALL | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **1** |
| Computers and Education: Artificial Intelligence | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** |
| Artificial Intelligence in Education: 20th International Conference Proceedings: Springer International Publishing | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| **Total:** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** | **1** | **4** | **1** | **8** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Venue** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **Total:** |
| Arab World English Journal (AWEJ)  | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| CALL and Complexity | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **1** |
| Journal of Research on Technology in Education | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** |
| Cogent Education | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** |
| Asia-Pacific Journal of Convergent Research Interchange | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| English Teaching | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **1** |
| Journal of Digital Convergence | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **1** |
| Korean Journal of English Language and Linguistics | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **1** |
| CALL communities and culture – short papers from EUROCALL 2016 | **-** | **-** | **-** | **-** | **-** | **-** | **2** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **2** |
| Advances in Human-Computer Interaction | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** |
| **Total:** | **-** | **-** | **1** | **-** | **-** | **-** | **2** | **1** | **1** | **2** | **1** | **1** | **-** | **2** | **11** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Venue** | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **Total:** |
| CALL in a climate of change: Adapting to turbulent global conditions–short papers from EuroCALL | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **-** | **-** | **-** | **1** |
| CALL and professionalization: Short papers from EUROCALL 2021 | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **1** |
| Intercultural Pragmatics | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| Computers & Education | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** |
| 22nd international conference on human-computer interaction with mobile devices and services | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| International Journal of Engineering Business Management | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **1** |
| Technology and the Psychology of Second Language Learners and Users | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **-** | **1** |
| **Total:** | **-** | **-** | **-** | **-** | **-** | **-** | **-** | **1** | **-** | **-** | **3** | **1** | **-** | **2** | **7** |
|  |

**Table 2. Overview of the review and analysis of the studies selected**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 1. Annamalai et al. (2023) | To observe students’ motivation to learn English via chatbots |  Twenty-five undergraduate students at a public university | A case study design: Semi-structured interviews | Duolingo, Mondly and Andy | Malaysia | Students had an enhanced sense of competence, autonomy, and relatedness, but some also felt using chatbots caused a lack of emotion.  |
| 2. Ayedoun et al. (2018) | The impact of adding the features of CSs and ABs to an AI agent on learners’ WTC | Forty Japanese undergraduate and graduate students  | Experimental design: Pre and post-tests (a WTC survey) | An intelligent personal agent created by the authors using the unity game engine, node.js, Wit.ai, and Dialogflow | Japan | The use of CS and AB strategies combined led to a higher level of WTC. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 3. Ayedoun et al. (2019) | To analyze WTC preferences of EFL learners while using an AI chatbot | Sixty university students | An experimental study: Pre and post-tests (a WTC survey) | An intelligent personal agent created by the authors | Japan | The CS + AB and CS versions was found more useful for more advanced level learners, while CS + AB and AB versions worked better for the lower level students. |
| 4. Ayedoun et al. (2020) | The use of ABs and CSs by students with different WTC levels  | Sixty university students | Experimental design: Pre and post-tests (a WTC survey); a system preference survey | Peter | Japan | The students with lower WTC favored CS+AB and AB strategies, whereas higher level WTC ones preferred CS+AB and CS versions.  |
| 5. Bao (2019) | To reduce the language learning anxiety of adult EFL learners through the use of a chatbot | Forty adult EFL learners (thirty-seven Thai and three Chinese) from college to doctoral level | A pilot study: Anxiety and AI usage surveys | Google Assistant | Thailand | Language learning anxiety of EFL learners was lowered after AI chatbot use for English language learning. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 6. Belda-Medina & Calvo-Ferrer (2022) | To explore the level of knowledge, satisfaction and perceptions as to the use of AI chatbots in English language learning among pre-service teachers | 176 undergraduate students | A sequential explanatory mixed methods research: Pre and post surveys  | Three AI chatbots (Replika, Kuki, and Wysa) | Spain (n= 115) and Poland (n= 61) | There were positive results in perceived usefulness, ease of use and attitudes about integrating AI chatbots, yet there was a moderate interest in future use. |
| 7. Chen et al. (2023) | The perceptions of students about Google Assistant regarding language learning | Twenty-nine EFL college students | A feasibility study: A questionnaire, a survey, and interviews | Google Assistant | Taiwan | Learners considered IPAs useful for pronunciation and vocabulary knowledge and reading, listening, and speaking skills enhancement.  |
| 8. Çakmak (2022) | To compare the L2 speaking performance and L2 speaking anxiety of students before and after the intervention | Ninety undergraduate EFLstudents | A mixed methods experimental study: An L2speaking anxiety scale, L2 speaking tests and a perception questionnaire | Replika | Turkey | The experimental group (EG) had a significantly higher speaking score, while L2 speaking anxiety was higher after the experiment. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 9. Dizon (2017) | The extent of Alexa’s capacity to comprehend L2 utterances and students’ opinions on its effectiveness | Four undergraduate students  | A case study: Learner-generated commands and interactive storytelling performance;interviews. | Alexa | Japan | Fifty percent of learner commands were understood by Alexa, while for interactive storytelling, the rate was 90%. |
| 10. Dizon & Tang (2019) | To examine the usefulness of a chatbot in promoting autonomous second language learning (ASLL) | Two ESL university students | A mixed method case-study design: Usage data and a survey | Alexa | Japan | While the students had positive attitudes, the use of chatbot did not contribute to ASLL. |
| 11. Dizon (2020) | To examine Alexa’s capacity to contribute to listening and speaking skills of EFL learners  | Twenty-eight first and second year undergraduate EFL students at an elective English-speaking course | A quasi-experimental design: Pre- and post-tests; listening and speaking tests; surveys | Alexa | Japan | There was no significant difference between the experimental group (EG) and the control group (CG) in listening scores, whereas speaking gains were significant. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 12. Dizon & Tang (2020) | EFL students’ perceptions towards IPAs for autonomous second language learning  | Fourteen Japanese university students | A mixed method case study: A survey  | Alexa | Japan | Although students showed an initial interest in using Alexa, nearly half of them abandoned its use, which would not indicate an autonomous learner behavior.  |
| 13. Ebadi & Amini (2022) | To explore the role of social presence and human-likeness stemming from the use of AI chatbots on learner motivation | 256 EFL learners at the departments of English Language Teaching, Translation Studies, and English Literature | A sequential explanatory mixed-methods design:Audio-recordings,   scales; a focus group interview | Computer  Simulation  in  Educational Communica-tion   (CSIEC) chatbot | Iran | According to SEM results, there was a positive relationship between human-likeness and learner motivation, and a positive and significant relationship between social presence and motivation.  |
| 14. El Shazly (2021) | The effect of the use of IPAs in lowering foreign language anxiety and developing speaking skills. | Thirty-eight undergraduate students | A quasi-experimental mixed methods design: Pre- and post- tests  | Web-based chatbots and Mondly | Egypt | FLA levels increased significantly after post-intervention, although there was a significant gain in post-intervention speaking scores. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 15. Ericsson et al. (2023a) | To look into cognitive, emotional, and social experiences of EFL learners in using the Embodied Conversational Assistant (ECA) in Enskill | Twenty-two seventh grade EFL students  | A digital logbook and a post-taskquestionnaire | Enskill English | Sweden | Learners were optimally challenged, engaged, and less stressed. Yet, about the social aspect, some regarded the ECA as “a deadpan machine”, while others attributed human qualities. |
| 16. Ericsson et al. (2023b) | To demonstrate the experiences of learners in speaking practice using virtual ECAs | Twenty-five seventh grade EFL students at a lower secondary school  | A feasibility study: System-generated metrics, questionnaires, logbooks, and interviews | Enskill English | Sweden | There were mixed experiences; that is, while some learners found ECAs fun and exciting, others reported that they had problems with voice recognition and intelligibility. |
| 17. Forsyth et al. (2019) | To analyze if the assessment of English language listening and speaking skills differ between a human and a chatbot assessor | Thirty-one 3rd and 5th grade students | A between subjects experimental design | TextEvaluator | the USA | The system was effective in measuring students’ language skills in comparison to a human. Yet, higher word count and the use of more complex language was determined in the human assessor condition. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 18. Fryer et al. (2017)  | The qualities of interaction with a chatbot and a human partner and their impact on task and sustained course interest  | A twelve-week compulsory university language course;122 first and second-year students | A longitudinal experimental study,three Likert-type scales for students (e.g., course and speaking task interest)  | Cleverbot | Japan | In chatbot condition there was a significant decline in task interest compared to human- human case. |
| 19. Fryer et al. (2019) | A follow up study aiming at determining if the decreased interest in chatbots continued over time.  | The same participants in Fryer et al., (2017) study were retested in the following semester. | Surveys and an open-ended questionnaire, and a listening and a reading test | Cleverbot | Japan | Decreased task interest with chatbot partners in Fryer et al. (2017) significantly increased. |
| 20. Fryer et al. (2020) | The relationship between self-efficacy, self-concept and task, domain, and course interest between chatbot and human partner scenarios | 128 first year university students | Latent modelling. Task interest, course interest, domain interest, course self-efficacy, and domain self-concept scales. | Chatbot employed is not specified. | Japan | While the chatbot condition predicted task interest, it did not predict course interest significantly. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 21. Han (2020) | To analyze the speaking skills development  | Forty-four freshmen students from a middle school | A pretest- treatment-posttest design | Alexa | Korea | The experimental group scored significantly higher in terms of speaking skills, vocabulary and grammar levels. Also, they had significantly lower levels of anxiety, and higher levels of language learning interest and motivation. |
| 22. Hassani et al. (2013) | To show the impact of the use of a chatbot on improving English language learners’ listening and speaking skills | Ten non-native graduate engineering students | A feasibility study | IVELL | Iran | There was a decrease in grammatical errors (3%) and pronunciation duration (16%) as well as an increase in proficiency level (11%). |
|  |  |  |  |  |  |  |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 23. Hsu (2022) | To investigate the attention and meditation levels, and the brainwave activities of EFL learners in Human-to-Human F2F, Human to Human Virtual, and Human to Chatbot Communication Environments | Thirty non-English major sophomore EFL learners  | Experimental neuroscientific design | An AI Chatbot, NeuroSky Mindwave, an EEG headset | Taiwan | In F2F condition attention was highest, whereas in chatbot condition meditation was the most prevalent. |
| 24. Hsu et al. (2023) | To analyze if there would be a statistically significant difference in the speaking scores of English language learners after interacting with a chatbot | 100 undergraduate students  | An experimental method | TOEIC Practice Chatbot (TPBOT) | Taiwan | The experimental group students had significantly higher oral test scores. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 25. Hwang et al. (2022) | To demonstrate the correlation between students’ learning behaviors and achievements and the impacts of use of SmartU English on learning outcome | Forty-three undergraduatestudents   | A  mixed methods experimental study: A paper test, an oral test, and interviews | Smart  UEnglish | Not Specified (NS) | EG students had significantly higher oral test achievement results as to lexical resource and pronunciation. |
| 26. Jeon (2022) | To demonstrate the views of students about the app in terms of satisfying their needs of autonomy, competence, and relatedness  | One hundred and twenty Korean EFL primary school students | A  mixed methods experimental study: Student  logs of app use,a SDT survey, and interviews | PengTalk | South Korea | The students who used the app continuously had significantly higher perceived competence and relatedness scores. |
| 27. Jeon (2024) | To illustrate facilitating and delimiting affordances of chatbots and their impact on learner motivation | Thirty-six Korean students at two primary classrooms  | A qualitative approach: Interviews and student-chatbot interaction logs | Unnamed chatbots created by the author via Dialogflow | South  Korea | Depending on users’ positive or negative experiences in using the chatbot, there were mixed effects on the learner motivation domain. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 28. Johnson (2019) | To test the efficiency of the Enskill speaking bot by collecting data from learners and to improve the application for future use | Intermediate level undergraduate students in Serbia (n=80) and Croatia (n=39) | Comparison of learner data based on cloud-based data collection, and 5-point Likert-scale and open-ended questions;108 CEFR A1/A2 simulations | Enskill English  | Serbia and Croatia | Learners reported increased proficiency in listening and speaking skills. |
| 29. Kim (2017) | To analyze the instances of negotiation of meaning between the student-student and student-chatbot groups | 123 undergraduate students  | An experimental design: Usage data and questionnaires | Indigo | Korea | There was a significant increase in the use of negotiation moves between the first and the last chat in the experimental group. |
| 30. Kim (2018) | To analyze if there is any difference in listening and reading scores between the two groups  | Forty-six freshman students  | Experimental design | Elbot | South Korea | The experimental group had a statistically higher increase in terms of listening skills, whereas there was no statistically significant difference as to reading test results. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 31. Kim et al. (2021) | To show how AI chatbots impact students’ speaking skills and motivation to use English language | Forty-nine undergraduate students | A pre- and post-test design: Speaking test scores and open-ended questions | Replika, Andy, and Google Assistant | Korea | Although they reported problems in intelligibility of learner utterances, there were statistically significant differences between the higher and the lower proficiency group as to fluency, intonation and stress in the “Read a text aloud" task and in all aspects in the “Respond to questions” task. |
| 32. Kim et al. (2022) | To assess the quality of the chatbot in terms of the suitability of language level,conversation continuity, and task performance achievement | 137Korean high school students | Not specified. | Ellie | South Korea | The mean of average CPS was 9.3, which was higher than the times in other traditional class tasks. |
| 33. Kwon et al. (2016) | To determine the efficiency of the developed chatbot by examining students’ experiences  | Twenty subjects with varying proficiency levels | A feasibility study | A task-oriented chatbot created by authors for English language learners | Korea | The average task success, turn success and turn length rates were 85.52%, 85.32% and 14.61 respectively. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 34. Lee & Jeon (2022) | To explore the viewpoints of primary school EFL students about a VCA aslanguage learning partners | Sixty-seven  Koreanprimary school EFL  students | A drawing task  and an interview | A VCA developed by authors and integrated into Google Assistant | South Korea | 71.6% of the students regarded the VCA as possessing human attributes. |
| 35. Lee & Lim (2023) | To verify if the developed chatbot capable of evaluating users’ pronunciation accuracy and sentence organization is effective for developing their speaking skills | A group of teachers and students (Particular characteristics were not specified) | A feasibility study | A chatbot created by authors for English language learners | Korea | The chatbot had a positive impact on students’ language skills by being a motivating source. |
| 36. Li et al. (2020) | To examine the effectiveness on a newly-developed task-based dialogue system | Twenty-eightundergraduate students | A system effectiveness study: User data and a questionnaire | A task-based dialogue system created by authors | Not specified | Many students complained about their utterances not being recognized by the system. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 37. Lin & Mubarok (2021) | To demonstrate the efficacy of using AI chatbots mind-mapping strategy on speaking skills development | Fifty undergraduate students  | Experimental design: An English oral test, the voice recordings of speech with the chatbot, chatbot memory | Replika  | Taiwan | The EG group employing the chatbot with mind-mapping strategy had significantly higher speaking scores. |
| 38. Morton & Jack (2010) | To carry out a cross-cultural empirical evaluation between the learners of English and French, who used an AI chatbot  | Twenty-eight students of French and forty-eight students of English at a junior high school | A usability study using a user centered design: Ability and motivation questionnaires and recognition accuracy rates | SPELL | China | There was an increased level of motivation among EFL learners, whereas French learners did not experience a similar increase. |
| 39. Morton et al. (2012) | To determine the efficiency of AI speech avatars | Forty-eight EFL students at a junior high school | Experimental design | A dialogue-based CALL system with virtual interactive game characters  | UK | There was a high level of engagement and grammar semantic value accuracy. Word-for-word recognition was low. The affective state of learners improved in time of the use. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 40. Moussalli & Cardoso (2016) | To illustrate how users exploit the chatbot system for improving their language skills | Four intermediate level English learners | A feasibility study | Alexa  | Canada | Learners viewed Echo as useful for improving vocabulary and pronunciation knowledge. The speech was intelligible although it was machine-like. |
| 41. Moussalli & Cardoso (2017) | To determine the efficiency of a chatbot in comprehending L2-accented speech | Eleven L2 participants from a variety of language backgrounds  | A feasibility study: Surveys and semi-structured interviews | Alexa  | Canada | Alexa had a high capability of comprehending learners' speech and to be understood by users. |
| 42. Moussalli & Cardoso (2019) | The impact of student-student voice-based and student-chatterbot communication on negotiation of meaning strategies and learners’ accentedness of speech | The same participants in Moussalli & Cardoso (2017) | A feasibility study: Surveys, judges’ rating on two pronunciation measures, and interviews | Alexa  | Canada | Rephrasing, changing the syntax, vocabulary and pronunciation were NfM strategies used. |
|  |  |  |  |  |  |  |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 43. Moussalli & Cardoso (2021) | To investigate if the use of a chatbot would help English language learners’ pronunciation of -ed allomorph as to phonological awareness, perception, and production | Eighteen undergraduate ESL students | A pretest-post-test mixed method design: Test outcomes and interview data | Alexa | Canada | The learners improved their phonological awareness of the -ed form. However, there were no significant differences for the perception and production. |
|  |  |  |  |  |  |  |
| 44. Park (2022) | To compare the groups chatbot-human in the VR world, human-human in the VR world, and traditional face to face peer groups’ L2 speaking gains | Thirty undergraduate EFL learners | An experimental study: L2 speaking tests | An AI chatbot in VR SPEAKIT (a speaking skills enhancement program) | South Korea | The chatbot-human group in the VR world had the highest mean increase among the three groups in terms of L2 speaking gains. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 45.Sydorenko, Daurio, et al. (2018)  | To enhance the pragmatic skills of English language learners through an AI chatbot | Twelve advanced level learners of English: Arabic (n=5), Burmese (n=1), Pashtu (n=1), Spanish (n=3),Turkish (n=1), and Vietnamese (n=1). | A qualitative case study: Interviews and simulation data | SimCon | The USA | Learners demonstrated increased pragmatic awareness. |
| 46.Sydorenko, Smits, et al. (2018)  | To analyze the development of oral communicative and pragmatic competence of L2 learners in the use of three chatbots | SimCon:Fourteen intermediate and eleven advanced learners HALEF: L2 learners with fifty-two different L1s in the Amazon Mechanical Turk platform | The design is not stated explicitly. Pre-test and post-test design; questionnaires with closed- and open-ended questions | HALEF and SimCon | The USA | Chatbots helped learners with a variety of backgrounds to improve communicative competence according to variables of learner proficiency levels with an emphasis on lower and higher order thinking skills. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 47. Tai & Chen (2020) | The effect of the use of Google Assistant on language learners’ WTC  | 112 eighth-grade Chinese EFL learners  | Experimental  | Google Assistant  | Taiwan | A significant increase in WTC, improved communicative performance, and lower speaking anxiety |
| 48. Tai & Chen (2022) | To analyze the effect of GA on students’ listening comprehension | Ninety-two ninth-grade EFLlearners | Experimental design: Listening comprehension tests,questionnaires and interviews | Google Assistant | Taiwan | There was a statistically significant difference in GA-Hub group in listening scores. |
|  |  |  |  |  |  |  |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 49. Timpe-Laughlin & Dombi (2020) | To explore the request strategies and modifications used by English language learners while using a chatbot | 107 conversations; on Amazon Mechanical Turk, i.e. Arabic (n=1), Bengali (n=2), Chinese (n=1), Filipino (n=2), Greek (n=1), Gujarati (n=1), Hindi (n= 12), Kannada (n= 1), Korean (n= 1), Malayalam (n= 7), Portuguese (n= 1), Russian (n=1), Slovak (n=1), Spanish (n=6), Tamil (n=5), Telugu (n=4), and Urdu (n=4) and Hungarian (n= 32) and Japanese (n=24) in the in situ data | A conversation analysis methodology | HALEF | Two universiti-es in Japan and Hungary, and Amazon Mecha-nical Turk’s platform | Direct requests were used the most, and among requests internal and want-statements and query preparatory were common.  |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 50. Timpe- Laughlin et al. (2022) | To explore ESL teachers’ views on spoken dialogue system (SDS) tasks | Sixteen ESL instructors at a university | Surveys, individual and focus group interviews | HALEF | United States | The teachers regarded SDS tasks positively, believing that they have potential in improving and assessing speaking skills.  |
| 51. Wang et al. (2022) | To investigate the social, cognitive, and teaching presences experienced by language learners after using an AI chatbot | 327 students at a primary school | AI usage data, learning outcomes, and attitudinal data | An AI chatbot created by authors | China | Cognitive and social presence and affection for the chatbot predicted the learning outcomes positively, whereas teaching presence had a negative impact. |
| 52. Wang et al. (2023) | To determine the variety of the different types of learners as to AI chatbot use | Sixteen EFL primary school students | Learners’ usage data and reflection essays | An AI chatbot created by authors | China | There were four clusters of students: Effective, passive, well-balanced and inefficient learners, each of which benefits from AI in their unique ways. |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 53. Wu et al. (2020) | To show the perceptions of native and non-native speakers of English in the use of a chatbot | Thirty-two participants, of whom sixteen were native English speakers and sixteen were Mandarin speakers | A feasibility study: A semi-structured interview | Google Assistant | UK | Non-native speakers of English believed that their linguistic needs were not met during the use of the chatbot due to issues such as IPA waking and turn taking problems. Whereas native speakers focused on more technical aspects such as the time gap between the production and perception.  |
| 54. Wu et al. (2023) | To determine users’ ratings of the created voice-based AI chatbot | Four undergraduate students | Focus groups and benchmarking | SpokenBot | Hong Kong | Students considered most traits as positive; yet, they rated NewsRead Aloud activity as not-effective. |
| 55. H. Yang et al. (2022) | To investigate the experiences of EFL students while using a task-based chatbot  | 314 EFL learners: 177 fifth or sixthgraders from elementary schools, and 137 first-year high schoolstudents | Conversation logs (student-chatbot speech) and questionnaires | Ellie | South Korea | There were high task success ratios in carrying out the three tasks (88.3% on average), although opinions regarding its use varied from neutral to positive. |
| **Study** | **Research Questions/Focus** | **Settings and Participants** | **Research Methodology** | **Technologies and/or Chatbots used** | **Countries** | **Main Findings/Results** |
| 56. C.T.Y. Yang et al. (2022) | To examine the listening and speaking development after using an AI chatbot | Thirty-four college EFL learners | An experimental design: Test scores and weekly evaluation forms | Google Assistant | Taiwan | There was a significant difference as to the speaking dimension, whereas there was no statistically significant difference in terms of listening skills. |
|  57. Zou et al. (2020) | To demonstrate students’ perspectives on opportunities and the challenges of the use of chatbots in practicing speaking in English language learning | 113 First and second year undergraduate EAP students | A sequential explanatory mixed methods design: A perception survey and interviews | Liulishuo, iFlytek and Chivox | China | Overall, students had positive perceptions about the integration of a chatbot in EAP curriculum. Yet, challenges such as poor voice recognition and intelligibility problems were reported. |
|  |  |  |  |  |  |  |

**Table 3. Theoretical backgrounds and focuses in the reviewed studies**

|  |  |  |
| --- | --- | --- |
| Theoretical background | *n* | Authors |
| Interaction Hypothesis  | 8 | Dizon (2017);Dizon (2020);Kim (2017);Morton & Jack (2010);Morton et al. (2012);Sydorenko, Daurio, et al. (2018);Tai & Chen (2022);Timpe-Laughlin et al. (2022) |
| Autonomous and self-access language learning | 8 | Dizon & Tang (2019);Moussalli & Cardoso (2020);C.T.Y.Yang et al. (2022) |
| Willingness to communicate (WTC) | 4 | Ayedoun et al. (2018);Ayedoun et al. (2019);Ayedoun et al. (2020);Tai & Chen (2020) |
| Foreign language anxiety | 3 | Bao (2019);Çakmak (2022);El Shazly (2021); |
| Interest development model | 3 | Fryer et al. (2017);Fryer et al. (2019);Fryer et al. (2020) |
| Skill acquisition theory | 2 | Sydorenko et al. (2018b); Timpe- Laughlin et al. (2022) |
| Self-determination theory | 3 | Annamalei et al. (2023);Jeon (2022);Morton & Jack (2010) |
| TAM and CHISM | 1 | Belda-Medina & Calvo-Ferrer, 2022 |
| Computers as Social Actors (CASA) | 1 | Ebadi & Amini (2022) |
| Theoretical background | *n* | Authors |
| The three factor theory of anthropomorphism | 1 | Lee & Jeon (2022) |
| Sociocultural theory | 1 | Ericsson et al. (2023b) |
| Constructivism | 1 | Park (2022) |
| Affordance theory | 1 | Jeon (2024) |
| The noticing hypothesis | 1 | Timpe-Laughlin & Dombi (2020) |
| Community of Inquiry (CoI) | 1 | Wang et al. (2023) |
| Theory of planned behavior | 1 | Wu et al. (2023) |
| Attention and meditation | 1 | Hsu (2022) |

## **Table 4. Strengths and challenges reported in the reviewed papers**

|  |  |  |
| --- | --- | --- |
| Strengths/Weaknesses | *n* | Reporting sample studies: Author(s) |
| *Strengths* |  |  |
| Developing language skills | 27 | Annamalai et al. (2023);Chen et al. (2023);Çakmak (2022);Dizon & Tang (2019);Dizon (2020);El Shazly (2021);Forsyth et al. (2019);Fryer et al. (2019);Han (2020);Hassani et al. (2013);Hsu et al. (2023);Hwang et al. (2022);Johnson (2019);Kim (2018);Kim et al. (2021);Lee & Lim (2023);Li et al. (2020);Lin & Mubarok (2021);Moussalli & Cardoso (2016);Moussalli & Cardoso (2021);Park (2022);Tai & Chen (2020);Tai & Chen (2022);Wang et al. (2023);H. Yang et al. (2022);C.T.Y. Yang et al. (2022);Zou et al. (2020) |
| Enhancing affective factors | 27 | Annamalai et al. (2023); Ayedoun et al. (2018);Ayedoun et al. (2019);Ayedoun et al. (2020);Bao (2019);Çakmak (2022);Dizon & Tang (2019);Ebadi & Amini (2022);El Shazly (2021);Ericsson et al. (2023a);Ericsson et al. (2023b);Fryer et al. (2017);  |
| Strengths/ Weaknesses | *n* | Reporting sample studies: Author(s) |
| *Strengths* |  |  |
| Enhancing affective factors |  | Fryer et al. (2019);Fryer et al. (2020);Han (2020);Hsu (2022);Jeon (2024);Kim et al. (2021);Lee & Jeon (2022);Morton & Jack (2010);Morton et al. (2012);Moussalli & Cardoso (2016);Moussalli & Cardoso (2021);Park (2022);Tai & Chen (2020);H. Yang et al. (2022);C.T.Y. Yang et al. (2022) |
| Facilitating communicative competence | 14 | Belda-Medina & Calvo-Ferrer (2022);Dizon (2017);Dizon & Tang (2019);Kim (2017);Kim et al. (2022);Kwon et al. (2016);Moussalli & Cardoso (2017);Moussalli & Cardoso (2019);Park (2022;)Tai & Chen (2020);Wang et al. (2022);Wang et al. (2023);Wu et al. (2020);Wu et al. (2023) |
| Increased pragmatic awareness/competence | 10 | Ayedoun et al. (2018);Ayedoun et al. (2019);Ayedoun et al. (2020);Dizon & Tang (2020);Ebadi & Amini (2022);Jeon (2024);Sydorenko, Daurio, et al. (2018); Sydorenko, Smits et al. (2018);Timpe-Laughlin & Dombi (2020)Tai & Chen (2020) |
| Strengths/ Weaknesses | *n* | Reporting sample studies: Author(s) |
| *Strengths* |  |  |
| Provision of a conversation partner/native speaker | 8 | Chen et al. (2023);Dizon (2017);Dizon & Tang (2019);Hwang et al. (2022);Jeon (2024);Park (2022);Tai & Chen (2020);Wang et al. (2023) |
| Promoting autonomy and self-regulated language learning | 5 | Annamalai et al. (2023);Dizon & Tang (2019);Dizon & Tang (2020);Jeon (2022);Jeon (2024) |
| Attentive to individual differences | 5 | Ayedoun et al. (2018);Ayedoun et al. (2019);Ayedoun et al. (2020);Lee & Lim (2023);Wang et al. (2023) |
| Increasing self-awareness of errors/indirect feedback  | 5 | Chen et al. (2023);Dizon (2017);Dizon & Tang (2019);Jeon (2024);Tai & Chen (2020) |
| Providing an immediate response | 4 | Belda-Medina & Calvo-Ferrer (2022);Jeon (2024);Tai & Chen (2020);Wang et al. (2023) |
| A source of authentic input | 3 | Chen et al. (2023);Jeon (2024);C.T.Y. Yang et al. (2022) |
| Rich multimedia content | 3 | Annamalai et al. (2023);Belda-Medina & Calvo-Ferrer (2022);Ebadi & Amini (2022) |
| Time flexibility | 2 | Annamalai et al. (2023);C.T.Y. Yang et al. (2022) |
| Strengths/ Weaknesses | *n* | Reporting sample studies: Author(s) |
| *Strengths*Space flexibility | 2 | Annamalai et al. (2023);Tai & Chen (2020) |
| *Weaknesses* |  |  |
| Problems with speech recognition and transcription and intelligibility of learner utterances | 15 | Belda-Medina & Calvo-Ferrer (2022);Dizon & Tang (2019);Dizon (2020);Ericsson et al. (2023a);Ericsson et al. (2023b);Fryer et al. (2019);Jeon (2024);Kim et al. (2021);Kim et al. (2022);Li et al. (2020);Morton et al. (2012);Tai & Chen (2020);Timpe-Laughlin & Dombi (2022);H. Yang et al. (2022);C.T.Y. Yang et al. (2022) |
| Unnaturalness of AI chatbot-human interaction | 9 | Annamalai et al. (2023)Chen et al. (2023);Ericsson et al. (2023a);Ericsson et al. (2023b);Jeon (2022);Jeon (2024)Lee & Jeon (2022);Moussalli & Cardoso (2016);C.T.Y. Yang et al. (2022) |
| Lack of explicit/corrective feedback | 5 | Chen et al. (2023);Hassani et al. (2013);Jeon (2022);Jeon (2024)Kim et al. (2021) |
| Network problems | 3 | Dizon & Tang (2020);Tai & Chen (2020);C.T.Y. Yang et al. (2022) |
| Linguistic restrictions of chatbots (e.g. low grammatical accuracy)  | 2 | Belda-Medina & Calvo-Ferrer (2022);Tai & Chen (2020) |
| Strengths/ Weaknesses | *n* | Reporting sample studies: Author(s) |
| *Weaknesses*Slowness to respond | 1 | Wu et al. (2020) |
| Too quick to respond | 1 | Tai & Chen (2022) |
| Long utterances | 1 | Tai & Chen (2022) |
| Chatbot’s proficiency level of the speech (e.g. too high) | 1 | Dizon & Tang (2019) |
| Overly sensitive to pronunciation errors | 1 | Chen et al. (2023) |
| Turn taking problems | 1 | Wu et al. (2020) |
| Lack of time | 1 | Kim et al. (2021) |
| A lack of multi-language support | 1 | Dizon (2017) |
| Error correction | 1 | Belda-Medina & Calvo-Ferrer (2022) |
| Data privacy | 1 | Belda-Medina & Calvo-Ferrer (2022) |
| Semantic coherence; maintenance of the conversation flow | 1 | Fryer et al. (2019) |

**Appendix.** **A list of the reviewed articles**

Annamalai, N., Eltahir, M. E., Zyoud, S. H., Soundrarajan, D., Zakarneh, B., & Al Salhi, N. R.

(2023). Exploring English language learning via Chabot: A case study from a self-determination theory perspective. *Computers and Education: Artificial Intelligence,* *5*, Article

100148. <https://doi.org/10.1016/j.caeai.2023.100148>

Ayedoun, E., Hayashi, Y., & Seta, K. (2018). 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*(1), 29–57.

<https://doi.org/10.1007/s40593-018-0171-6>

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). L2 learners’ preferences of dialogue agents: A key

to achieve adaptive motivational support? In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, R. Luckin (Eds.). *Artificial Intelligence in Education: 20th International Conference*, (pp. 19-23). Springer International Publishing.

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

Bao, M. (2019). Can home use of speech-enabled artificial intelligence mitigate foreign language

anxiety–investigation of a concept. *Arab World English Journal*, *5*, 28-40.

<https://dx.doi.org/10.24093/awej/call5.3>

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in

language learning. *Applied Sciences*, *12*(17), Article 8427. <https://doi.org/10.3390/app12178427>

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. <https://doi.org/10.1080/10494820.2020.1833043>

Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL*, *16*(2), 113-131.

Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *TESOL Journal*, *8*(4), 811–830. <https://doi.org/10.1002/tesj.353>

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>

Dizon, G., & Tang, D. (2019). A pilot study of Alexa for autonomous second language learning. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 107-112). Research-publishing.net.

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *JALT CALL Journal*, *16*(2), 107-120. https://doi.org/10.29140/jaltcall.v16n2.273

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>

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking

performance: A case study. *Expert Systems*, *38*(3), Article e12667. <https://doi.org/10.1111/exsy.12667>

Ericsson, E., Lundin, J., & Sofkova Hashemi, S. (2023a). From deadpan machine to relating socially: Middle school students’ experiences speaking English with embodied conversational agents. *Journal of Research on Technology in Education*, 1-17. <https://doi.org/10.1080/15391523.2023.2213902>

Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023b). Fun and frustrating: Students’ perspectives on practicing speaking English with virtual humans. *Cogent Education*, *10*(1), Article 2170088. <https://doi.org/10.1080/2331186X.2023.2170088>

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., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning

experiences, interest and competence. *Computers in Human Behavior*, *93*, 279–289. <https://doi.org/10.1016/j.chb.2018.12.023>

Fryer, L. K., Thompson, A., Nakao, K., Howarth, M., & Gallacher, A. (2020). Supporting self-efficacy beliefs and interest as educational inputs and outcomes: Framing AI and human partnered task experiences. *Learning and Individual Differences*, *80*, Article 101850. <https://doi.org/10.1016/j.lindif.2020.101850>

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*(7),71-80.<https://doi.org/10.47116/apjcri.2020.07.07>

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>

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. 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. <https://doi.org/10.1080/10494820.2021.1960864>

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. Advance online publication. <https://doi.org/10.1080/09588221.2022.2095406>

Jeon, J. (2022). 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. (2024). 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>

Johnson, W. L. (2019). Data-driven development and evaluation of Enskill English. *International Journal of Artificial Intelligence in Education*, *29*(3), 425–457. <https://doi.org/10.1007/s40593-019-00182-2>

Kim, N.Y. (2017). Effects of types of voice-based chat on EFL students’ negotiation of meaning according to proficiency levels. *English Teaching*, *72*(1), 159-181. <https://doi.org/10.15858/engtea.72.1.201703.159>

Kim, N. Y. (2018). A Study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence*, *16*(8), 19-26. https:/doi.org/10.14400/JDC.2018.16.8.019

Kim, H. S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students’ communication skills. *Korean Journal of English Language and Linguistics*, *21*, 712-734. <https://doi.org/10.15738/kjell.21..202108.712>

Kim, H., Yang, H., Shin, D., & Lee, J. H. (2022). Design principles and architecture of a second

language learning chatbot. *Language Learning & Technology*, *26*(1), 1-18. <http://hdl.handle.net/10125/73463>

Kwon, O.-W., Kim, Y.-K., & Lee, Y. (2016). Task-oriented spoken dialog system for second-language learning. In S. Papadima-Sophocleous, L. Bradley & S. Thouësny (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 237-242). Research-publishing.net.

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*, 1-26. Advance online publication. <https://doi.org/10.1080/09588221.2022.2067182>

Lee, K. A., & Lim, S. B. (2023). Designing a leveled conversational teachable agent for English

language learners. *Applied Sciences*, *13*(11), Article 6541. <https://doi.org/10.3390/app13116541>

Li, K. C., Chang, M., & Wu, K. H. (2020). Developing a task-based dialogue system for English language learning. *Education Sciences*, *10*(11), 306. <https://doi.org/10.3390/educsci10110306>

Lin, C. J., & Mubarok, H. (2021). Learning analytics for investigating the mind map guided AI

chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society*, *24*(4), 16-35. <https://doi.org/10.2307/48629242>

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>

Morton, H., Gunson, N., & Jack, M. (2012). Interactive language learning through speech-enabled virtual scenarios. *Advances in Human-Computer Interaction*, *2012*, Article 389523. <https://doi.org/10.1155/2012/389523>

Moussalli, S., & Cardoso, W. (2016). Are commercial ‘personal robots’ ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley & S. Thouësny (Eds.), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325-329). Research-publishing.net.

Moussalli, S., & Cardoso, W. (2017). Can you understand me? Speaking robots and accented speech. In K. Borthwick, L. Bradley & S. Thouësny (Eds.), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 217-221). Research-publishing.net.

Moussalli, S., & Cardoso, W. (2019). 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>

Moussalli, S., & Cardoso, W. (2021). Intelligent personal assistants and L2 pronunciation development: Focus on English past -ed. In N. Zoghlami, C. Brudermann, C. Sarré, M. Grosbois, L. Bradley, & S. Thouësny (Eds.), *CALL and professionalisation: Short papers from EUROCALL 2021* (pp. 226-231). Research-publishing.net.

Park, H. (2022). Effects of virtual reality-based English learning on Korean university students’ speaking ability. *Multimedia-Assisted Language Learning*, *25*(4), 93-119. https:/doi.org/10.15702/mall.2022.25.4.93

Sydorenko, T., Daurio, P., & L. Thorne, S. (2018). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, *31*, 157-180. <https://doi.org/10.1080/09588221.2017.1394326>

Sydorenko, T., Smits, T. F. H., Evanini, K., & Ramanarayanan, V. (2018). Simulated speaking environments for language learning: Insights from three cases. *Computer Assisted Language Learning*, *32*(1-2), 17–48. <https://doi.org/10.1080/09588221.2018.1466811>

Tai, T. Y., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners’

willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485-1502. <https://doi.org/10.1080/10494820.2020.1841801>

Tai, T. Y., & Chen, H. H. J. (2022). The impact of intelligent personal assistants on adolescent EFL learners’ listening comprehension. *Computer Assisted Language Learning*. Advance online article. <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>

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 article. <https://doi.org/10.1080/09588221.2022.2056203>

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in

AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, *194*, Article 104703. <https://doi.org/10.1016/j.compedu.2022.104703>

Wu, Y., Rough, D., Bleakley, A., Edwards, J., Cooney, O., Doyle, P. R., Clark, L., & Cowan, B. R. (2020, October). See what I’m saying? Comparing intelligent personal assistant use for native and non-native language speakers. In *MobileHCI ‘20: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–9). ACM. <https://doi.org/10.1145/3379503.3403563>

Wu, C. H., Lam, H. Y., Kong, A., & Wong, W. L. H. (2023). The design and evaluation of a digital learning-based English chatbot as an online self-learning method. *International Journal of Engineering Business Management*, *15*, 1-22. <https://doi.org/10.1177/18479790231176372>

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, *34*(3), 1-17. <https://doi.org/10.1017/S0958344022000039>

Yang, C. T. Y., Lai, S. L., & Chen, H. H. J. (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>

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). Artificial intelligence technology for EAP speaking skills: Student perceptions of opportunities and challenges. In M. R. Freiermuth, & N. Zarrinabadi (Eds.), *Technology and the psychology of second language learners and users* (pp. 433–463). Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-34212-8>