The Effect of Multimodal Conversational AI on Job Interview Anxiety and Performance among ESL Students

Amany Haroun Alkhayat

Submitted in partial fulfillment of the

requirements for the Degree of Doctor of Education in

Teachers College, Columbia University

2026

Abstract

The Effect of Multimodal Conversational AI on Job Interview

Anxiety and Performance among ESL Students

Amany Alkhayat

Speaking a second language is often considered one of the most anxiety-inducing skills to acquire. For many ESL learners, limited vocabulary, fear of judgement, and lack of sufficient practice in a crowded classroom can impede their speaking performance. Although the literature has predominantly focused on interventions in general English classrooms, English for Specific Purposes courses, where English focuses on professional communication, remain underdeveloped, with most research focused on writing and reading. This gap is evident in the context of job interviews, which require not only linguistic knowledge but also familiarity with domain-specific vocabulary and discourse competence. This study employed a quasi-experimental, mixed-method design, where experimental groups practiced job interview questions with a multimodal embodied conversational AI (MECAI) in mixed reality (MR) simulation (Group1), or ChatGPT voice application (Group 2). The control group did not receive any treatment and relied on traditional classroom activities instead. Students (N = 55), ages 18-64 years from a community college participated in the study. All students completed a pretest and were randomly assigned to three conditions. The experimental conditions were assigned to a six-week intervention. After the intervention, all students completed a posttest questionnaire. Finally, 36 students were interviewed to explore their perceptions of the study intervention. The pre-and posttest measures comprised eight mock job interview questions, where students' responses were evaluated through a rubric that included seven criteria: fluency, pronunciation, intonation and stress, grammar and sentence structure, vocabulary use, content relevance, and the use of work-based examples. The pretest also included 33 Foreign Language Anxiety Scale (FLCAS) items rated on a 5-point Likert scale and heart rate measures during the pre-and post-mock job interviews. The interaction with AI conversation logs negotiation of meaning (NoM) patterns within the Task-based Language Teaching method (TBLT) were examined using a Computer

Mediated Discourse Analysis (CMDA) coding scheme. The reflection interviews consisted of 11 questions examining students' perceptions of their conditions, the teacher's role, experimental students' interaction with AI and the traditional condition experience, and whether each condition contributed to anxiety reduction and improved speaking performance. Key findings showed that although there was no significant difference between all groups in the pre-intervention mock job interview, FLCAS, or heart rate, both AI groups (MECAI in MR and ChatGPT voice app) showed significant post-intervention improvement, with the MR group demonstrating the most improvement in fluency, vocabulary, and total language performance. However, despite the fact that students in the AI conditions expressed anxiety reduction during reflection interviews, posttests of FLCAS and heart rate demonstrated no statistically significant reduction in physiological or self-reported anxiety levels. The experimental groups conversational logs showed more NoM pattern production than the ChatGPT voice group. Both AI groups helped address gaps in teachers and curriculum knowledge in ESP-oriented pedagogy, where speaking remains underexplored.

Table of Contents

List of Tables

List of Figures and Equations

Acknowledgements

My journey in the doctoral program at Teachers College was shaped by countless people. For many years, this journey has been truly transformative, where technology, learning, and human connection have been intertwined. Insights from great professors, colleagues, and impactful peers from research and industry conferences, and above all, my family, have added meaning to my growth as a human, scholar, and practitioner.

From the very beginning, my children: Lojain, Mohammed, and Omar, were part of every stage of this journey. Their patience and understanding through late nights, laughter through rough times, and pure, unconditional love impart true meaning to this work.

My deepest gratitude goes to the Technology, Learning, and Media (TML) Department, whose vision continues to include, enlighten, and inspire student researchers. My deepest gratitude goes to my advisor, Dr. Sandra Okita, for her guidance and forward-thinking vision that I witnessed through participating in her lab and reading her work. Her mentorship not only shaped my approach as a researcher but also guided the direction of this study. I am also grateful to Dr. Erik Voss, whose encouraging words during the AIRiAL conference strengthened my self-confidence, and whose Natural Language Processing course inspired research collaborations with other researchers and shaped the way I approach computational linguistics. I deeply appreciate the opportunity to work with Dr. Amra Sabi-El-Rayess, who welcomed me as a VR graduate research assistant on her Homeland Security-funded Belonging project. An inspirational project that impacted many, young and old, and made educators ponder belonging and educational displacement in schools.

Outside TC, I am truly indebted to Dr. Jonathon Richter, the visionary and influential president of the Immersive Learning Research Network. Since 2020, I have felt like I belong to a great institution, where I continued my life-long learning journey, collaborating with and inviting many professors and thought leaders under Dr. Richter's leadership.

I sincerely admire the Graduate Student Life and Development (GSLD) program for encouraging me to establish and become president of the TC student organization, Intelligent Reality in Education Association, the first EdTech student organization at TC devoted to AI in education. My deepest appreciation to Alana Crosby, Emy Yamamoto, and Payan Medina for providing me with support, patience and guidance.

I also wish to express my sincere appreciation to Professor Michelle Vera and Professor Ernest Bartledes at Hudson County Community College, who were excited by this study and its potential and made this study possible. I am also thankful to Professor Kevin Park, the Department Chair at the New York Institute of Technology and The HIVE research lab, for their support in my data collection phase and for providing me with the necessary head-mounted displays (HMPre-and I needed for my study.

Being Awarded the TML Microgrant 2025, I'd also like to express my appreciation to Dr. Jin Kuwata, Dr. Jooeun Shim, and Ms. Catherine Kinkead for their guidance and support. I am also grateful to my colleague Mengqian Wu, who collaborated with me on the grant proposal and workshop. I deeply appreciate the Wonda AI team: Arnaud Dressen and Ashley Etemadi for supporting the AI-Powered Learning Simulation Workshop.

Finally, every community I collaborated with and every individual I learned from shaped this dissertation and the individual I have become today. To my family and all those who supported me, this study is a testament to the impact you have had on me.

AA

Chapter I

Introduction

Speaking a second language is widely perceived as the most important skill for ESL students (Richards, 2008). It is defined as "a productive aural/oral process that consists of using grammatical rules, cohesive devices, lexical items, phonological rules for expressing one's thoughts and feelings in speech" (Hammad & Ghali, 2015, p. 53). An ESL student's ability to clearly articulate their ideas clearly and engage in meaningful conversations is essential for academic success, career readiness, and effective interpersonal communication (Byram & Morgan, 2014; Crystal, 2008; Byram & Morgan, 2014; Crystal, 2008). However, developing this skill presents significant challenges. A learner's success is frequently measured by their ability to carry out a conversation in the target language (Nunan, 1991). The complex interplay of producing correct pronunciation, appropriate vocabulary, fluency, and grammatical accuracy can provoke Foreign Language Anxiety (FLA) and impede speaking performance. Accordingly, FLA, often stemming from the fear of making mistakes, can directly hinder classroom participation, slow speaking progress, and ultimately limit the learner's ability to communicate effectively in real-world personal and professional scenarios (MacIntyre, 2007). Consequently, FLA emerges as a central barrier to ESL learners' speaking development (Gosiewska-Turek, 2018; Amengual-Pizarro, 2018; Woodrow, 2006, Marzec-Stawrarska, 2015).

In their seminal work on FLA, Horwitz et al. (1986) distinguished between general trait anxiety and FLA. They describe FLA as a situation-specific form of anxiety in which a learner reacts to the complexity and intricacy of ESL classroom demands, shaped by their negative internal perceptions, emotional responses, and behavior. To investigate FLA, Horwitz et al. developed the Foreign Language Classroom Anxiety Scale (FLCAS), which has been widely used in Second Language Acquisition (SLA) research to measure learners' anxiety levels. Research using the Horwitz Foreign Language Anxiety Scale (FLCAS) identified a significant negative correlation between FLA and speaking improvement.

Types of Anxiety

Anxiety is commonly identified in the psychological literature as a psychological construct rather than a reaction to a distinctly specific stimulus. Hilgard et al. described it as "a state of apprehension, a vague fear that is only indirectly associated with an object" (1971, cited in Scovel, 1991, p. 18).

However, in language learning research, anxiety is widely perceived as having an impact on performance (Horwitz, 2001). Psychologists typically distinguish between three categories of anxiety: trait, state, and situation-specific anxiety.

Trait anxiety refers to a person's relatively stable personality traits. Researchers define trait anxiety as "stable individual differences in a tendency to respond with an increase in state anxiety while anticipating a threatening situation" (Tovilović et al., 2009, p. 492). Trait anxiety reflects a general predisposition to anxiety across many situations rather than a temporary situation-specific type of anxiety (Scovel, 1978, cited in Ellis, 1994).

In contrast, state anxiety is a type of anxiety that arises as an immediate reaction to a situational, temporary, specific condition, such as taking an important test. Once the triggering situation ends, anxiety diminishes (Tovilović et al., 2009).

The third category is situation-specific anxiety. It refers to anxiety that constantly appears in particular situations rather than across all situations. It is recurring and multifaceted but confined to specific types of situations, such as public speaking or classroom participation. In language learning contexts, this type of anxiety is especially pertinent, as it is triggered repeatedly by communicative situations that demand performance (Ellis, 1994). As speaking anxiety is repeatedly triggered by communicative demands, research has focused on identifying the underlying factors that contribute to its development.

Causes of Speaking Anxiety

Research has identified several causes of speaking anxiety. Students' low speaPre-andoficiency levels, which may be challenging for them in the classroom, can cause anxiety (Sparks & Ganschow, 1991). Traditional classroom environments in terms of teaching style, type of activities, and classroom size can also trigger anxiety (Jackson, 2002; Oh, 1992; Oxford, 1999a; Powell, 1991; Samimy, 1989; Spielmann& Radnofsky, 2001). Individual factors that relate to the student such as age, their own beliefs about learning and learning strategies can also hinder speaking performance (e.g., Bailey, Daley, & Onwuegbuzie, 1999; Brownetal, 1996; Campbell, 1999; Dewaele, 2002; Ehrman& Oxford, 19Pre-anddner, Day, &MacIntyre, 1992; Gardner, Smythe, & Brunet, 1977; Gregersen& Horwitz, 2002; Oxford, 1999b, cited in Andreda & Williams 2006). Horwitz et al. (1986) identified three main factors that cause anxiety. These are Communication Apprehension, Fear of Negative Evaluation, and Test Anxiety. Other researchers, such as Young (1991), suggested six causes of anxiety: learners' beliefs about anxiety, teachers' beliefs about teaching S/FL, learners' beliefs about S/FL learning, assessment, classroom activities, and interactions between students and teachers.

The causes of speaking anxiety appear to stem from an intricate interaction between learners' internal beliefs and the external classroom environment, which directly impacts their language performance. This connection has led researchers to investigate the effects of anxiety on speaking performance.

Effects of Anxiety on Speaking Performance

For many years, the link between anxiety and language learning has drawn the attention of researchers because of its possible impact on language learning performance. Despite the fact that studies since the 1990s have shown high and inconsistent correlations between anxiety and language proficiency, recent meta-analyses provide strong evidence that FLA has negative effects on ESL students' overall achievement. Different meta-analyses show that the relationship between L2 anxiety and language achievement indicates a moderate-to-strong negative correlation of $r = -.36$ across 105 independent samples (Teimouri et al., 2019), with a subsequent study on FLA in English as a Foreign Language contexts finding an even stronger negative effect size of $r = -.61$ (Dickman, 2021)). These findings provide evidence of the negative role of anxiety, particularly in oral performance settings, where speaking anxiety has been identified as a significant moderator of second language achievement (Dickman, 2021). The literature also identifies that the type of skill is a significant moderator of achievement, indicating that listening and speaking anxiety has the most negative effect on performance (Dickmen, 2021). Research on the correlation between FLA and speaking performance reveals that as speaking anxiety increases, speaking performance decreases (Manda & Irawati, 2021; Rachmawati, 2020).

Research on university ESL students has demonstrated significant levels of anxiety, where 50% of all ESL students in higher education experience high degrees of anxiety that can affect them physically, psychologically, or socially (Höl & Kasımi, 2022). Bodily symptoms of anxiety can include erratic heartbeats and profuse sweating. The psychological impact of anxiety can cause fear of judgement or criticism. As for the social aspect, learners may be reluctant to communicate in the classroom. All these factors contribute to challenges in language comprehension and poor speaking performance (Kondo & Ling, 2004; Höl & Kasımi, 2022).

Prior research employing interventions to reduce ESL students' anxiety and improve their speaking performance has integrated technological tools such as virtual reality and conversational artificial intelligence (Dizon & Tang, 2019; Pertiwi & Kusumaningrum, 2021), holographic teachers (Cerezo et al., 2019), and game-based language learning (York, 2019). While the findings show positive student perceptions of these interventions and a reduction in anxiety, they also reveal a gradual decline in students' interaction with the intervention due to a lack of a specific pedagogical framework that guides the interaction (Dizon & Tang, 2019), a focus on pronunciation only (Cerezo et al., 2019), or the exclusion of meaningful interactions.

Moreover, the literature has identified several methodological flaws and conceptual limitations in research that has employed interventions to improve language learning performance and reduce anxiety (Elahemer & Said, 2022). These limitations include small sample size (Nazligulan et al. 2019; Stupar-

Rutenfrans et al. (2017) and Zacarin et al. (2019) conducted short-term experiments (Wang et al., 2020; Dincer et al., 2020) without control groups (Ramamuruthy, 2019), lacked personalized feedback when employing conversational agents (Wang et al., 2020), made unjustified intervention selections, had students with limited speaking competency, and used interventions that heightened anxiety (Altunkaya, 2018; El Shazly, 2021).

Despite the challenges and limitations of previous research, strategies that can lower students' anxiety have been suggested. These strategies depend on integrating technologies that can provide ample personalized speaking opportunities and student-centered approaches (Cantos et al., 2024). Research also shows that teaching methods such as Communicative Language Teaching (CLT), which emphasizes meaningful rather than accurate language use (Richards & Rodgers, 2001), and the use of real-world tasks that enhance language practice–an approach known as Task-Based Language Teaching (TBLT)–are widely endorsed (Willis & Willis, 2007). TBLT is a widely researched teaching method recognized for its potential to foster both fluency and accuracy (Nunan, 2004). Simultaneously, technology integration in teaching speaking is perceived as a powerful opportunity to improve students' speaking skills and reduce their anxiety (Sharma, 2024). There has been extensive research on the integration of digital tools such as video conferencing, online language exchange, game-based language learning, and other forms of computer-mediated communication (CMC), where real-life scenarios are incorporated and provide authentic interaction with peers and native speakers (Chapelle, 2003; Warschauer & Healey, 1998). The incorporation of these approaches is usually referred to as technology-mediated TBLT, which aims to increase students' speaking performance and reduce their anxiety (Ziegler, 2016).

Problem Statement

Despite the limitations mentioned in the previous sections, a significant gap persists between the potential of these interventions and their practical outcomes, associated with instructional environment challenges such as large class sizes in the traditional classroom that prevent personalized learning and interaction, and limited resources; learner-centered challenges such as FLA that impede students' speaking performance; and instructor-focused challenges such as the lack of specific teacher expertise in teaching speaking, especially in contexts like job interviews (Brown, 2007; Burns, 2019). Furthermore, the claim that technology reduces learner anxiety often relies on self-reported data, creating a scarcity of objective empirical evidence regarding state anxiety in technology-mediated interactions (Baralt & Gurzynski-Weiss, 2011). Another challenge is that a predominant focus within the literature has been skewed toward general computer-mediated communication (CMC) general ESL speaking interactions in which everyday conversational English is taught in contrast with research on English for Specific Purposes (ESP) where high stakes, context-dependent speaking tasks demand more than general conversational fluency. This area of research remains underdeveloped in the current body of research (Kim & Namkung, 2024) especially for learners with diverse backgrounds and proficiency levels (Chong & Reinders, 2020). Research often disregards anxiety related to career content itself, where students may have anxiety due to their lack of knowledge of the specific professional field in English, not only the language itself (Ayuningtyas et al., 2022).

Accordingly, the literature emphasizes the critical need for research that examines not only how technology fosters speaking performance and reduces anxiety, but also how learners use the technology during the task to better understand the specific mechanisms by which technology helps make learning happen (Ziegler, 2016). Research shows that ESL learners' interactions with technology can be significantly beneficial to their L2 development, such as negotiation for meaning and modified output, which successfully occurs during interactions within technology-mediated environments (González-Lloret, 2003; Lee, 2001). However, the literature indicates that the quantity and quality of these interactions are highly mediated by the tool and L2 task employed. For instance, studies have shown that meaning negotiation is affected by task type, with one-way information-gap tasks being highly effective (Blake, 2000) and predominantly triggered by lexical rather than grammatical items (Pellettieri, 2000; Tudini, 2003). Moreover, different modalities, such as voice chat and text chat, elicit unique patterns of repair and negotiation, highlighting the need to investigate the specific learner-technology process (Jepson, 2005; Yanguas, 2010). Currently, the majority of TBLT efficacy research has been conducted in written text-chat environments, leaving the interactional dynamics of oral, multimodal, and immersive contexts, which are highly relevant for speaking practice, relatively underexplored (Ziegler, 2016). Therefore, future research must move beyond general performance gains to explore the pedagogical efficacy of emerging technology by conducting multi-methodological investigations explored by students before and after the intervention, analyzing the actual patterns of interaction and negotiation, and facilitating the empirical confirmation of its effectiveness in reducing speaking anxiety and improving speaking skills.

Purpose Statement

To address challenges in speaking anxiety, particularly those related to the traditional classroom environment, limited authentic materials, and limitations in instructors' knowledge and curriculum, this study proposes the integration of two artificial intelligence techniques: multimodal embodied conversational AI (MECAI) in Mixed Reality (MR) and ChatGPT voice application.

To provide a basis for this argument, a pilot study with ten participants demonstrated that immersive multimodal embodied AI can enhance social presence and NoM patterns and can support a more positive perception of anxiety. Students in both experimental groups practiced job interview questions for six weeks. Both groups used NoM patterns to negotiate meaning with AI (See Table 1), but the MECAI in MR group showed the highest NoM patterns. Students in the traditional conditions, on the other hand, used their first language to ask their peers when there was a communication breakdown during instruction and tended to memorize responses instead of asking questions that related to their job experience or experimenting with language for fear of judgement. In their reflection interviews, the MECAI in MR group expressed a desire to acquire a Head-Mounted Display for more personalized, immersive practice and showed a stronger preference for AI tools that simulate real-life scenarios through gestures, eye tracking, and natural conversation. Students in the ChatGPT group also expressed positive views on using the ChatGPT voice application. They enumerated many advantages, including ubiquity, ease of use, and complementing teachers' knowledge. On the other hand, students in the traditional condition, who relied on peer roleplay and often resorted to their first language, showed limited engagement and tended to memorize interview responses rather than developing authentic communicative competence.

Table 1

Frequency of NoM Strategies between MECAI in MR Group and ChatGPT voice application Group

Interaction Type	**MECAI**	**ChatGPT**
Confirmation/Comprehension Checks	32	18
Clarification Requests	28	12
Repetition	20	10
Rephrasing/Correction	18	7

Recent advances in large language models (LLMs) and immersive technologies have facilitated their integration into instructional technology and language learning research. Moreover, developments in embodied conversational AI, LLMs, lip-syncing techniques, gaze, and gestures allow MECAI agents to mimic human conversational behavior and provide authentic, contextualized, and situated experiences (Kukulska-Hulme & Shield, 2018; McGivney, 2025). The MECAI in MR can foster social presence with a pedagogical agent (Oh et al., 2018; Han & Bailenson, 2024) that can support an anxiety-free environment.

MR was chosen over virtual reality (VR) for several reasons: 1) MR creates an immersive, yet non-isolating environment, where learners are aware of their physical environments while engaging with the MECAI in MR (Stretton et al., 2018) ; 2) MR, compared to VR, does not induce motion sickness, providing an invisible interface (Weiser, 1999); Chen & Duh, 2019 where students can focus on the conversational task. Despite the growing interest in AI-enhanced immersive learning, there is limited research in the language acquisition research on how this technology can improve speaking performance and reduce anxiety especially in context-specific fields such as ESP.

Therefore, this study aimed to examine how integrating embodied AI agents in mixed reality and voice-based ChatGPT mobile applications within a TBLT framework can enhance speaking performance and reduce speaking anxiety among English learners. This study contributes to research on AI-mediated language learning by bridging advances in embodied technologies with affective and performance outcomes in ESP speaking contexts.

Approach

This quasi-experimental, mixed-method design study examines the potential of integrating multimodal conversational AI interventions to reduce speaking anxiety and improve speaking performance in pre-intermediate-level English learners in the context of mock job interviews. This study will investigate whether practicing with MECAI in an MR environment or with a voice-based ChatGPT mobile application can result in measurable improvements in learners' speaking performance and a reduction in their foreign anxiety levels compared with a control group that does not receive any intervention.

The experimental groups will engage in an instructional intervention for six weeks, whereas the control group will receive only traditional classroom instruction. The first experimental group will interact with MECAI in an MR environment that mimics real-life job interview scenarios. The second experimental group will use the voice-based ChatGPT mobile application. Both interventions enable learners to practice structured interview responses to prepare for answering eight specific job interview questions in a naturalistic spoken dialogue format.

All groups will complete a pre- and posttest battery that includes 1) mock job interviews that are video recorded and evaluated using a rubric that assesses accuracy, fluency, and whether answers show personal relevance rather than generic or memorized lists; 2) The Foreign Language Classroom Anxiety Scale (FLCAS) self-report questionnaire to measure their perceived anxiety; and 3) Heart Rate measures using Polar H1+ monitor during the mock job interviews to measure their physiological anxiety levels.

Following the six-week intervention, participants will be interviewed to reflect on their experience with their assigned conditions and their perception of the use of conversational AI tools versus the traditional method. The interviews will be audio recorded, transcribed, and thematically analyzed.

This study attempts to fill a gap in research related to speaking anxiety in English for Specific purposes, an area that remains critically underdeveloped. By focusing on this under-researched topic and specific professional language needs, this study fills a gap in the literature.

Research Questions:

To accomplish this, the following research questions were examined:

1- How does interacting with an AI agent in Mixed Reality, the ChatGPT voice app, or traditional methods impact participants' overall speaking performance across seven rubric dimensions (Fluency; Pronunciation; Intonation and; Stress; Grammar and; Sentence Structure; Vocabulary Use; Content and; Relevance of Answers; Use of Work-Based Examples) from pre- to post-mock job interview, and which interaction modality yields the best overall speaking performance?

2- How do interactions with an AI agent in mixed reality, the ChatGPT voice app, and traditional methods impact participants' self-reported anxiety levels and heart rates during the pre- and post-mock job interviews with an external examiner?

3- What are the different strategies used by students to negotiate meaning when interacting with AI agents in mixed reality or using the ChatGPT voice app?

4- What are the perceptions of the students in the experimental and control groups about the effectiveness of these conditions in helping them overcome their anxiety and improve their speaking performance?

Anticipated Outcomes

It is expected that all three groups (MECAI in MR, ChatGPT voice app, and traditional group) will demonstrate improvement in mock job interview scores, show less anxiety in the FLCAS posttest, and have a low heart rate. However, I hypothesize that the first treatment condition, where students practice job interview questions in the MECAI in MR, will show the greatest gains. This is because research has shown that conversational AI in immersive environments can improve speaking performance. This is due to the immersive and embodied affordances of MR and embodied conversational AI that foster presence, realism, and interaction, which promote communicative competence (Deng & Tavares, 2021; Lee, 2022). Participants in this group are also expected to show lower anxiety levels in the post-FLCAS and heart rate. Social presence and a judgement free immersive environment have been shown to reduce psychological arousal and anxiety in prior research (Bailenson, 2021; Pan et al., 2024; Zheng & Cheng, 2018).

Participants in the ChatGPT voice-based app group are expected to show moderate gains in speaking performance and smaller physiological changes in FLCAS and heart rate. This may be due to disembodiment. To explain, voice-based chatbots provide low-pressure, more accessible environments that can encourage continuous use and experimentation with language (Le et al., 2022; Azizimajd, 2023);

however, they lack multimodal cues such as gaze and gestures that can enhance engagement and social presence (Jung et al., 2019; Kim et al., 2021).

The traditional condition was expected to show the smallest improvement in the posttest mock job interview and the highest anxiety indicators in both the FLCAS and heart rate. This is due to reliance on traditional role-play activities, lack of curriculum-related authentic materials that can help them improve their professional skills, and their dependence on memorization rather than practicing how their responses should relate to their professional experiences (MacIntyre & Gardner, 1994; Payne & Whitney, 2002). However, the presence of an external examiner is expected to challenge all conditions, elicit evaluative stress, and may elevate heart rate and amplify anxiety during the mock job interview posttest (Abad et al., 2021; Mathobela et al., 2024).

In terms of identifying the negotiation of meaning patterns during the six-week intervention with the two conversational AI conditions, it is anticipated that the MECAI in the MR group will demonstrate more negotiation of meaning patterns than the ChatGPT voice-based app group. This is due to the embodied nature of embodied conversational AI, which fosters presence (Lim et al., 2025). Existing literature has not yet explored NoM patterns with ECAI in immersive environments; however, a few studies have examined text-based and voice-based interactions and traditional role-play activities, which showed that disembodied conversational AI significantly increased more negotiation of meaning patterns by students compared to student-student traditional activities (Kim, 2017; Kim & Lee, 2023; Shim, 2007). Participants using voice-based chats with chatterbots, in the literature, employed strategies such as confirmation checks and repetition to repair communication breakdowns in contrast to the student-student communication. It is also anticipated that the traditional condition would employ code-switching, a strategy students use by using their first language to ask their peers questions to resolve communication breakdowns, which can affect speaking performance (Færch & Kasper, 1983; Yi & Sun, 2013).

In terms of the reflection interviews following the posttest measures, it was anticipated that both experimental groups would have positive opinions about the entire experiment. The traditional condition will also have a positive attitude; however, in the pilot study, the control students felt challenged as they did not use AI. In their reflection interviews, they expressed interest in using AI to complement teachers' knowledge of the subject.

In summary, it is hypothesized that the MECAI in the MR condition will have the highest gains in terms of the mock job interview posttest, less anxiety in terms of the FLCAS and heart rate, and will also display more meaning negotiation patterns than the other groups. The overall experience among the groups can have a positive impact on all groups.

Rationale and Significance

The significance of this study stems from its potential to provide evidence that supports new pedagogies in second language teaching, especially in research related to ESP and employing LLMs in educational technology (Lopez-Gazpio, 2025) that can supplement teacher knowledge in areas that require diverse professional knowledge (Pang, He, & Marlow, 2025). Moreover, MR as an educational tool used in professional training and simulations (Sun et al., 2023; AIT Austrian Institute of Technology, 2024) provides an immersive interactive environment that can foster situated learning and the feeling of immersion.

The integration of embodied AI agents in mixed reality offers an innovative way to enhance language learning by providing immersive and interactive practice that can simulate real-life communication scenarios. ECAs in immersive environments create authentic learning spaces that provide situated practice and contextualized learning experiences (Kukulska-Hulme & Shield, 2018; McGivney, 2025). The MECAI functions as a pedagogical agent capable of fostering social presence and the feeling of "being with" another (Oh et al., 2018; Han & Bailenson, 2024). This sense of social presence, along with a safe, Judgement-free immersive environment, has been shown to effectively reduce learner anxiety (Pan et al., 2024).

This study also employed objective and subjective measures of anxiety. Measuring heart rate and students' anxiety self-reports using FLCAS during pre- and post-mock job interviews can provide a comprehensive picture of what the students experience before and after the intervention. Most research in language learning exploring speaking anxiety have mainly focused on self-reports which can be subject to bias or inaccurate self-perception (Shachter, Kangas, Sweller, & Stewart, 2022; Sosa-López & Mora, 2022). Physiological measures, such as heart rate, provide objective data on anxiety responses, which enhances the validity of affective measurement and can offer insights into how anxiety can impact speaking performance in ESP contexts.

Additionally, the negotiation of meaning analysis can provide rich data on interactional strategies when communication breakdown occurs when students interact with AI and how this contributes to learners' speaking performances. Research shows that in the traditional classroom, negotiation of meaning was always initiated by the teacher 94% of the time when the teacher felt a communication breakdown, whereas in computer-mediated environments, students became more active as 36% of negotiation patterns were initiated by the students (Shim,2007). Moreover, the qualitative data from the reflection interviews can offer rich insights into the students' reflections on task authenticity, assigned condition, teacher role, AI role in improving their speaking performance, and reducing their anxiety levels.

Beyond the classroom and students' benefits, the significance of this study extends to industry and commercial applications. There are many commercial VR and augmented reality applications that target language learning, but they may not focus on professional, contextualized high-stakes needs, such as job interviews for non-native speakers. Some of these applications use conversational AI, which may digress from the main topic or have rule-based, limited interactions. Most applications that use conversational AI in immersive environments face challenges in adapting to diverse users' linguistic and cultural backgrounds, providing context-aware feedback, and maintaining their engagement over time (Sikström et al., 2024; Han & Ryu, 2025). By investigating adaptive MECAI in mixed reality for professional ESP tasks, this study can inform the design of more effective training tools for corporate, healthcare, and service industries in which employees must communicate in high-stakes, cross-cultural situations. The evidence from this study can guide developers in enhancing personalization, social presence, and pre- and posttest-based learning, which are critical for both learner motivation and real-world skill transfer (Bian & Zhou, 2022; Ke, Lee, & Xu, 2016).

This study also explored the CMDA approach for examining the negotiation of NoM in immersive environments compared with ChatGPT voice-based applications. Most research on language learning has focused on the negotiation of meaning in traditional classroom interactions or computer-mediated communication (CMC) settings, such as text chat (Blake, 2000; Smith, 2003; Sauro, 2009), video conferencing (Yanguas, 2010; Satar & Özdener, 2008), or interacting with mobile app chatbots (Yin & Satar, 2020), where learners engage in clarification requests, confirmation checks, and reformulations to resolve communication breakdowns (Foster & Ohta, 2005; Smith, 2003). It is important to understand how intelligent immersive environments that facilitate social presence can elicit negotiation episodes and how this differs from voice-based environments. This can inform the design of AI-enhanced immersive environments and how this method of interaction can promote authentic and communicative experiences.

The Researcher

I am a full-time doctoral student in the Technology, Media, and Learning (TML) Program at Teachers College (TC). I first joined the program as an MA student in 2019 with strong enthusiasm and passion for integrating technology into the classroom and exploring the impact of chatbots on students' learning. Courses in my program shifted my perspective from a technology-centric approach to designing inclusive learning experiences that cater to diverse students' needs.

Joining Dr. Okita's Gizmo EdTech lab was a transformative experience and the first step in understanding empirical research. This experience drove the use of heart rate monitors in the present study. My colleagues at the Gizmo EdTech lab: Dr. Eddie Lin, Dr. Elliot Hu-Au, Dr. You Zhang, Dr. Nicola Law, and Marcus Cheung, provided a seamless experience for me that made my doctoral journey deeply rewarding. Moreover. Joining Dr. Voss' research group and studying his course on Natural Language Processing empowered my research trajectory, which I applied to a research paper that I later published at an IEEE conference.

Immersive learning has been a passion of mine since my research agenda focused on conversational AI in immersive learning environments. This research focus led me to take several positions in different institutions and collaborate on research with several colleagues, including teaching Artificial Intelligence in immersive learning environments at the Media University of Applied Design in Germany and teaching human–computer interaction, advanced UX research, and UX design for VR at the New York Institute of Technology. In addition, I served as a course assistant for MSTU 5000, taught by Dr. Joey Lee, where I gained research insights from him and helped students understand immersive learning. This experience culminated in receiving the Microgrant award for my proposal on AI-Powered learning simulations, with my colleague, Mengqian Wu, providing support throughout the project development.

My experience beyond TC was highly transformative and rewarding. I was awarded the AI Curricular Innovation Grant at Hudson County Community College in 2025, the Most Engaging App Award from Ringling College in 2024, and other awards. I served as a committee chair and moderator for many conferences, such as the Immersive Learning Research Network, which has always been my home under Dr. Jonathan Richter's leadership, Gamicon conference, Women in Voice, TESOL International Electronic Village, and New York State TESOL. These experiences solidified my interest in exploring how conversational AI and intelligent technologies can foster equitable, engaging, and personalized learning environments.

Definitions

ChatGPT: A Large Language Model (LLM) that stands for Generative Pre-trained Transformer (GPT). It generates human-like text and has text-to-speech and speech-to-text features (Wu et al., 2023).

EFL (English as a Foreign Language): A field of language instruction in which English is taught in a non-English speaking country (Longcope, 2009).

Embodied Conversational AI: A branch of artificial intelligence that focuses on creating a virtual character that can communicate with the learner through a flat interface via video or an immersive 3D avatar. MECAI can use gestures and facial expressions and interact with humans through text or voice. This avatar can be used in AI simulations and intelligent tutoring systems (Yang et al., 2025).

ESP (English for Specific Purposes): An approach to teaching English that focuses on professional contexts, academic disciplines, and linguistic needs (Coffey, 1984).

ESL (English as a Second Language): A field of language instruction taught in English-speaking countries (Longcope, 2009).

Head-Mounted Display (HMD): A wearable technology that includes sensors, a camera, and is connected with a controller that can enable users to interact with others or objects in an immersive environment (Renganayagalu, 2021).

Mixed Reality (MR): An immersive technology that enables users wearing an HMD to see virtual objects blended into their physical environment (Milgram & Kishino, 1994).

Presence: The subjective and mental sense of “being there” in a virtual environment (Wilkinson et al., 2021, September).

Prompt Engineering: The process of designing input prompts that guide LLMs to generate a desired output (Schulhoff et al., 2024).

SDK (Software Development Kit): A set of development tools and documentation that enable developers to create certain applications (Sharma, 2019).

Virtual Reality (VR): A 3D computer-generated environment that can be accessed on a laptop or through HMD (Milgram & Kishino, 1994; Slater, 2009).

Chapter II

Literature Review

Introduction

To understand the context and challenges involved in developing speaking skills in language learning, several areas of research must be explored. This literature review first examines the challenges that impede speaking performance and induce anxiety in learners. It then examines the trajectory of prior research that explored interventions in general ESL and ESP classrooms, underscoring the pedagogical and technological approaches employed to reduce anxiety and improve students' speaking skills. The literature review provides examples of interventions, such as the use of computer-mediated communication, mobile-assisted language learning, and chatbots, both prior to and following the advent of LLMs. The review further focuses on speaking skills in specific contexts, particularly ESP, and underscores the underdevelopment of research in this area. Finally, the literature presents the affordances of conversational AI and immersive learning and Mixed Reality, providing a solution to support ESP learners in specialized contexts.

Speaking Skill and Barriers to Learning

To communicate orally in a second language, ESL learners must first acquire the fundamental skill of speaking. When teachers prepare for a speaking lesson, the lesson objective is to provide students with the topics and classroom experience they need to develop communicative competence (CC). Communicative competence is a skill that enables learners to express their needs, wants, and ideas (Abdiganieva & Shamuratov, 2022). Among these objectives is the capacity to comprehend and interact with people in a second language (Krashen & Terrell, 1983). Therefore, practicing speaking in class is crucial for learners to develop their fluency and accuracy in order to communicate in real-life scenarios (Burns, 2012).

Studies have revealed that most college students learning a second language experience significant anxiety when it comes to practicing speaking in front of the class (Ansari, 2015; Idrus &

Saleh, 2007; Savasci, 2013). In the US, there are many ESL classes with students from different backgrounds and tongues, which can make classroom interaction difficult. The fear of being judged or trolled is a common fear among students (Pinatih, 2021; Arham et al., 2016; Krebt, 2017; Ben Tahir & Hanapi, 2017).

Research on speaking performance has shown that there are major factors that can prevent students from developing their speaking skills. The most cited challenges include affective factors such as fear of Judgement, whether teased by classmates or corrected by teachers, making mistakes, and anxiety (Ahmed, 2018; Kaur, 2022; Thao & Nguyet, 2019; Tuan & Mai, 2015), limited vocabulary (Paneerselvam & Mohamad, 2019), language barriers, and cultural differences (Ferris & Tagg, 2019), limited exposure to the language due to classroom size (Rajendran & Yunus, 2021), and traditional teaching methods (Mthembu & Pillay, 2014) that may lead to limited opportunities for students to practice speaking effectively (Liew & Aziz, 2022). Research shows that students who struggle with these elements usually become passive learners in the classroom (Paneerselvam & Mohamad, 2019).

Speaking and General ESL for Academic Purposes

An overview of the scholarly landscape that focuses on interventions to improve speaking performance and reduce anxiety primarily falls within the literature on general English courses or English for Academic purposes courses, where the focus is solely on everyday English or practicing oral skills for public speaking. Early interventions in general ESL classrooms focused on teaching basic listening skills and gradually evolved into employing technologies that are more sophisticated and immersive. On the other hand, when it comes to the literature regarding job interviews, we find that there are two trajectories: research that solely focuses on native speakers and research that is underdeveloped and attempts to address speaking challenges among learners who take English for specific purposes that focus on career readiness and workplace success. To understand the gaps in research that address speaking challenges in the area of job interviews, it is important to trace the trajectory of interventions across related fields: general ESL courses, ESP, and job interviews for native speakers. We can determine whether the existing approaches and interventions can be adapted or whether new approaches and interventions are needed for such a specialized topic.

Interventions in General ESL Classrooms

Research on speaking anxiety has grown immensely over the past few decades. This area of research was influenced by Krashen's (1981) "affective filter" hypothesis, a theoretical barrier that can hinder language acquisition due to negative emotions and anxiety. Later seminal work by Horwitz, Horwitz, and Cope (1986) indicated that foreign language anxiety (FLA) is a unique concept that can be differentiated from general anxiety. They suggested that FLA comprises three elements: text anxiety, speaking inhibition, and fear of judgement. This was corroborated by findings that showed that anxiety in language learning classrooms was often higher than in other academic courses (Muchnick & Wolfe, 1982; MacIntyre & Gardner, 1991).

Subsequently, research conducted between the 1990s and the 2000s validated the negative correlation between anxiety and language performance, particularly in speaking tasks (MacIntyre & Gardner, 1991b; Phillips, 1992). These findings show the adverse effects of anxiety on students' vocabulary acquisition (McIntyre & Gardner, 1991), which cause mental blocks (Chen & Chang, 2004) and lead students to hesitate to participate in classroom speaking activities for fear of judgement (Öztürk, 2003; Tsiplakides & Keramida, 2009).

Consequently, interventions for ESL learners enrolled in general English courses can revolve around two areas: pedagogical and technological. These approaches aimed to reduce anxiety while simultaneously improving speaking performance at the same time. Early interventions introduced a judgement-free classroom environment (Young, 1991) using drama techniques (Sevim, 2014). More recently a significant body of research has focused on technology-enhanced language learning interventions such as computer-mediated Communication (CMC) including virtual worlds and virtual and augmented reality, Mobile-Assisted Language Learning, and AI and chatbots,

Computer-Mediated Communication (CMC)

CMC refers to "any communication pattern mediated through the computer" (Metz, 1994, p.33). The rise of computers in education has sparked the interest of researchers, and a plethora of research studies have been conducted in the area of language learning comparing the use of CMC with traditional classrooms. CMC can be divided into two types according to timing, temporality, and feedback immediacy: asynchronous, meaning communication that does not happen in real time, and synchronous, meaning communication that happens in real time. In his book, An Invitation to CALL, Hubbard (2021) introduces another dimension to synchronous communication, namely quasi-synchronous communication, which means that although learners interact in real time through text or voice messages, they have the chance to modify the text before hitting the send button, which gives them more opportunities and less anxiety than real-time communications that require immediate speaking production.

Speaking skill has been the most extensively examined skill in language learning research, given its importance and difficulty (Mahdi, 2014). Tracing the research trajectory on using CMC to improve speaking skills, research on CMC shows that text-based chat CMC reduces speaking anxiety (Abrams, 2003; Tudini, 2007; Warschauer, 1996), but when it comes to oral CMC, the effect is not the same. (Satar & Özdener, 2008). This is due to the lack of paralinguistic cues, namely gestures, which may lead to “awkward silence” (Hampel & Baber, 2003). Other reasons can be the time the learner takes to respond through voice chats and the learner’s perception of the task difficulty (Sauro, 2011). Research has also found that when students do not feel rushed and can edit their voice messages before sending them, anxiety lessons (Poza, 2005). This may not mimic real-life authentic situations where students feel pressured to respond. When it comes to specialized speaking courses that tackle career readiness for students from diverse professional backgrounds, there is little research on English for professional or vocational purposes, where the language is more specialized and teachers do not have the required knowledge for these courses. Previous research attempted to address this issue through language proficiency training, AI-driven activities, and cognitive behavioral therapy has not fully addressed this issue due to their limited scope and limited real-time adaptability (Powell et al., 2018; Hosseini et al., 2024).

Other Interventions

Altunkaya (2018) used activity-based oral presentation classes based on a constructivist approach. Verbal and non-verbal cues were taught to the students, and they engaged in various activities. There was no control group. Pre- and posttests for anxiety and performance were conducted. However, learners' anxiety increased because they were asked to speak in front of an audience despite their interest in participating in the activity. This aligns with Kana's (2015) findings on pre-service teachers increasing anxiety when they desire to participate in activities due to personal expectations. This also aligns with Aydin's (1999) research, which had the same findings. Others, such as Sevim (2014), found that the use of some drama techniques for non-native speaking pre-service teachers through creative writing and performance alleviated anxiety. Chen (2018) used a 3D holographic platform through an interactive learning system where students use gestures to interact in a second language learning environment. The designed learning environment decreased the learners' anxiety.

Elshazly (2021) used virtual reality to reduce ESL learners' anxiety. The intervention slightly increased learners' anxiety. This could be due to learners being in a classroom setting and the potential for failing the course. Additionally, there was also no control group in ElShazly's study. Second, students might have struggled with VR technology given that the design of Mondly lacks cues that can help learners navigate the VR environment easily. Han and Keskin used mobile apps, such as WhatsApp, to reduce anxiety, and found it to be effective. This was due to students' familiarity with the app, self-assessment, feedback provided by others and the teacher, and positive self-awareness due to engagement with the activity. However, the students expressed doubt about using WhatsApp in a classroom setting, where they preferred a traditional classroom setting. WhatsApp is not sufficient to replace teachers in some activities.

Hashim et al. (2019) used Flipgrid as an intervention for all skill levels. Speaking anxiety decreased, and students' confidence increased. Kilic et al. (2018) used psychoeducation techniques such as cognitive behavioral therapy to reduce students' anxiety. It was effective and reduced students' anxiety, but the sample size was small at a single institution, which may affect generalizability. There is a need to explore the integration of such techniques with teaching practices and pedagogies of ESL. Lee and Kim (2018) explored the effect of input (receptive) and output (productive) based planning on Korean students' anxiety where students focused on interpersonal strategies for communication. Although the study focused on low-proficiency learners, there were differences in students' abilities. The difference between the input and output planning groups was not significant.

Nazligul et al. (2019) combined psychoeducation strategies using cognitive behavioral therapy with virtual reality exposure intervention. This was effective in reducing anxiety. However, the sample size was small, and self-report measures were used. Pontillas (2020) used Popsispeak, an app that provides teacher feedback. The intervention decreased the students' anxiety. While no significant relationship was found between public speaking anxiety and oral communication skills, the intervention resulted in a noticeable reduction in public speaking anxiety and an enhancement in oral communication skills among the students. The combination of practical activities, continuous reinforcement, and motivation provided by the teacher-researcher played a crucial role in achieving these results. The Pontillas study did not find a statistically significant correlation between reduced anxiety and improved communication skills, which might suggest that other factors are also at play in how these skills and anxieties develop or are mitigated. The scope of this study was confined to a specific context and a small number of students.

Ramamuruthy (2019) used a task-based approach to reduce students' anxiety levels. The researcher used sheets adapted from Richard's speaking activities. These activities are divided into three types: interactions, transactions, and performances. However, this study focused on ESL learners with low proficiency. The type of task that reduces anxiety should be explored. Stupar-Rutenfrans et al. (2017) explored the use of mobile virtual reality. The intervention was efficient for students with high anxiety. Wang et al. (2020) used conversational agents, Alexa, to reduce learners' anxiety. The intervention helped learners reduce their anxiety. This study did not include a control group. The tutoring sessions lacked personalized content and instruction, potentially limiting the effectiveness and user engagement of the intervention.

Zacarin (2019) employed virtual reality and cognitive behavioral therapy. Participants showed improved self-assessment scores in public speaking, indicating enhanced speaking confidence and reduced anxiety. Despite the high stimulus of the VRE, participants reported a general reduction in anxiety during public speaking situations, both immediately after the sessions and in follow-up scenarios. However, there was a small sample size and therapist bias, and participants noted some immersion-breaking issues with the VR environment, such as poor graphic quality and unrealistic avatar movements, which could have affected their anxiety levels.

Despite the increased interest from CMC researchers in Computer-Assisted Language Learning (Kessler, 2018), there are still gaps in research on how CMC can reduce anxiety and improve ESL speaking performance. Ziegler (2016) recently highlighted the need for research that examines the extent that technology might affect learners' anxiety, as well as their subsequent L2 development and performance.

In the context of job interviews as social situations, anxiety during job interviews can adversely impact social behavior and job interview performance (Carver, 1979; Carver & Scheier, 1981). The candidate's negative self-perceptions negatively influence job candidates' social behavior in terms of effectively engaging in a social situation, such as an interview, and interpreting interviewers' responses. Heerey and Kring (2007) found that anxious job candidates tend to overlook social cues, such as smiling, when the interviewer smiles. This results in a negative perception from the interviewer that the interviewee is cold and non-assertive.

Several factors contribute to job interview anxiety, including the cognitive demands of interview questions and social interactions inherent in the process. Research has listed the causes of job interview anxiety that impact job interview performance. First, interview questions are tasks that involve cognitive demands and social interactions, as candidate responses are evaluated by another person. This results in anxiety exerting a substantial influence on job candidates, impairing their cognitive performance and social situations, and negatively influencing interviewers' perceptions. Eysenk et al. (2007) proposed the attentional control theory of anxiety, which posits that job candidates tend to focus on their worries (i.e., self-negative thoughts) instead of their performance, which distracts them from doing well in job interviews.

Research exploring the differences between native and non-native speakers faces many challenges in gatekeeping encounters that impact their job interview performance (Gumperz, 1992; Kerekes, 2006, 2007). These challenges include language barriers, such as fluency and idiomatic expressions, which make it difficult to convey their qualifications and personality effectively (Gumperz, 1982; Roberts et al.,1992). Another issue is the cultural differences in terms of communication styles and interview norms across cultures, which can lead to misunderstandings. For example, non-native speakers might miss subtle cues because of differing expectations. These causes increase anxiety, which adversely affects the cognitive function and performance. Interlocutors belonging to the same first language (L1) background are viewed as more congruent and hence have fewer chances of miscommunication than interlocutors with different cultural and linguistic backgrounds (Qutub, 2014).

Research has found that excessive speaking anxiety can lead to performance impairment, loss of concentration, and reticent behavior (Abbasi, Khalil, & John, 2019). Many suggestions, such as adopting a humanistic learner-centered approach, improving speaking proficiency through training, and creating a more friendly and low-anxiety learning environment, have been made to reduce speaking and foreign language anxiety (Galti, 2017; Sinaga et al., 2020; Young, 1991). More recently, another line of speaking anxiety research has investigated the effectiveness of various techniques continues to develop. However, there is a lack of consensus regarding which learning elements can be changed to manage speaking anxiety.

Employment interviews are one of the most commonly used methods for assessing the suitability of job candidates. Because they play such a prominent role in the hiring process and involve a social situation in which one is evaluated by a stranger, employment interviews provoke anxiety in many candidates (McCarthy & Goffin, 2004).

Efforts to reduce job interview anxiety have included various types of interventions. For example, emotional strategies and breathing therapy have been effective in reducing anxiety among nursing students (Dincer et al., 2020). Virtual reality (VR) has also been explored as a tool for reducing anxiety, although the results have been mixed. While some studies have found VR interventions to be effective, others have noted increased anxiety due to the unfamiliarity of the technology or the lack of supportive cues (Elshazly, 2021).

Speaking and the literature on Computer Assisted Language Learning

CALL is an approach to language learning that focuses on the use of technology to facilitate language learning (Hubbard, 2017). As defined by Heift and Schulze (2015), "Computer-assisted language learning (CALL) contributes to individual practice in language learning through Tutorial CALL, Intelligent CALL, and individualization through tailored learning sequences and contingent guidance." The first stage of CALL was influenced by Thorndike (1913), Pavlov (1927), and Skinner (1974), and was called Behaviorist CALL (Warshauer, 1996).

Behaviorist CALL

The foundation of instruction in behaviorist CALL was based on repetitive drills and cloze tests (Warshauer & Healey, 1998). As a result, numerous CALL tutoring systems were developed in the 1960s, including PLATO (Programmed Logic for Automatic Teaching Operations). It was created for use with the current mainframe systems. According to Bitzer et al. (1961), the PLATO system comprised vocabulary drills, short grammatical lessons and exercises, and periodic translation tests. However, in the 1970s, the introduction of microcomputers and criticism of the behaviorist approach from theoretical and pedagogical angles led to the rejection of the behaviorist approach and ushered in a new phase called **Communicative CALL.**

Communicative CALL and Communicative Language Teaching

Communicative Language Teaching (CLT) emanated from the criticism of traditional teaching methods in language learning, which focused on grammar and structure. Rather than emphasizing linguistic competence, CLT focuses on communicative competence, which emphasizes the capacity to utilize language effectively in ordinary social contexts. Therefore, understanding what is meant by communicative competence is essential to comprehending communicative language teaching. It refers to the ability to negotiate meaning, understand sociolinguistic and grammatical principles, and determine the appropriateness of a given statement (Savignon, 1972, 1983, 1987). CLT promotes second language learning by prioritizing content over form, providing comprehensible input (i.e., language that is at the learner's level of proficiency), fusing various discourse types, and offering methods and activities that facilitate language teaching (Benati, 2009).

Communicative CALL is an extension of CLT, where technology is incorporated into the classroom. According to Warshauer (1996), this stage in CALL emerged after criticisms of Behaviorist CALL and traditional teaching methods. He lists several uses of technology in the classroom where the focus is on form instead of structure, grammar is not taught explicitly, speaking skill teaching methods rely on motivating students to use authentic language, and constructive feedback is provided. With this method, students may use their language skills in a technology-mediated setting by completing language activities through software interactions (Michel & Lehuen, 2002).

This phase includes several applications that have been developed and utilized in various computing roles. While some programs functioned as tools, others were more like computer stimuli for the students. This literature review focuses on the function of computers as tutors, which was introduced by communicative CALL. In these applications, the computer was programmed to provide the correct answer. However, “the process of finding the right answer involves a fair amount of student choice, control, and interaction.” (Warshauer, 1996, p.1).

Intelligent CALL (ICALL)

Intelligent Computer-Assisted Language Learning involves the use of computers and artificial intelligence in language instruction that incorporates Natural Language Processing (Gamper & Knapp 2002). The ICALL field emerged from CALL when some artificial intelligence applications proved to be sufficiently developed for language learning. These applications include grammar checkers, intelligent tutoring, spelling checkers, and others. With the advancement of research on Automated Speech Recognition (ASR), ICALL supports the teaching of speaking skills, including pronunciation. Gramper and Knapp (2002) provide several types and techniques of ICALL in their review paper. The first is expert systems, which are programmed to provide students with the required information, such as verb forms and spelling. The system also provides feedback based on stored knowledge of students’ possible mistakes. The second is intelligent tutoring systems (ITSs). The system has four modules: expert, learner, tutor, and user interface. The expert module consists of knowledge that should be learned. The learner module has knowledge of the learner’s proficiency level and needs. The tutor module includes interactions, strategies, and learning objectives. The graphical user interface allows the learner to interact with the system.

Other systems include adaptive systems in which the questions can change depending on how the learner responds to them. Natural language processing is also used for grammar and spelling checkers. Natural language generation is a more advanced field of language learning that encompasses semantics and pragmatics. With the help of ASR, these technologies can allow “users to have extensive” face to face” dialogues in real time with virtual characters” (p.7).

CALL Approaches to Enhancing ESL Speaking Skills

According to research, there are strategies that encourage students to practice speaking and develop their communicative competence while creating a stress-free and anxiety-reducing environment in the classroom. Debates have been used in some traditional classroom settings to help students improve their speaking and critical thinking abilities (Zare & Othman, 2014). In other studies, students' speaking abilities were improved through games like guessing games, sharing, and caring, which improved their oral performance (León & Cely, 2010). Some researchers have used social media to enhance their students' speaking abilities by using Instagram posts to discuss field trip memories and other stories related to learners' personal experiences (Handayani, 2016, May; Mansor & Rahim, 2017).

Studies have explored enhancing ESL students' speaking skills through different approaches. These approaches include employing effective teaching strategies and using technology. For example, some studies have highlighted the importance of student-centered classrooms in terms of nuanced teaching methods and learning environments that are anxiety-free and promote practice-oriented tasks.

Language Learning Chatbots

Chatbots are computer programs that use text and speech to converse with humans. Chatbots have been used in a variety of online environments, both commercial and entertaining, but one potential application that has received recent attention is their use in second language learning (Coniam, 2014), specifically in Communicative Intelligent Computer-Assisted Language Learning (CommICALL) (Wilske, 2014). Research suggests that chatbots have the potential to help ESL/EFL students overcome learning problems by facilitating learner engagement and learnability, improving productive skills and grammar, facilitating positive feedback, and inhibiting affective filters. Freyer and Carpenter (2006) claim that chatbots can "provide a means of language practice for students anytime and virtually anywhere" (p. 8).

When it comes to interactional competence, chatbots have the potential to be a cheap and easily accessible substitute for human interaction in the field of second language learning (Bibauw et al., 2022). Because of their availability (Huang et al., 2018), chatbots improve linguistic engagement among learners (Shawar & Atwell, 2007; Lee et al., 2022). Interaction with chatbots boosts students' willingness to communicate (Peng & Woodrow, 2010).

In order to understand the research on chatbots, there is a need to address the literature before Large Language Models (LLMs) and the literature after LLMs. The purpose of this study is to show how chatbots were used in different speaking activities, research limitations and gaps in research, how LLMs solved some of these problems, and identify remaining areas in need of further investigation and development within this domain.

In this literature review, I divided chatbot research into the era before LLMs and the era after LLMs because of the major breakthrough that LLMs have contributed to chatbots and language learning.

Chatbots in the Pre-LLM Era

Previous studies have shown controversial views on chatbots. According to Fryer and Carpenter (2006), chatbots are capable of providing opportunities for learners, such as reducing anxiety, offering endless interaction opportunities with the learner, allowing learners to interact with them using text or speech recognition to practice speaking or writing, motivating learners to practice the second language, and including error corrections. However, Jia (2004) and Chantarotwong (2005) doubted the ability of chatbots to assist in language learning. They found that chatbots have limitations in terms of lack of memory and context awareness, expected responses, and lifeless character. However, human conversations can have the same attributes at certain times (Paltridge, 2007). Nevertheless, when it comes to pedagogy, research shows that chatbots Despite significant advancements over earlier conversational systems, chatbots appear to have much progress to make before accomplishing a goal in terms of pedagogy (Coniam, 2008a, 2008b, 2014; Williams and Compernolle, 2009), and more research is needed to explore this potential before they can be used effectively in SLA (Fryer and Carpenter, 2006).

Previous studies have shown controversial views on chatbots. According to Fryer and Carpenter (2006), chatbots are capable of providing opportunities for learners, such as reducing anxiety, offering endless interaction opportunities with the learner, allowing learners to interact with them using text or speech recognition to practice speaking or writing, motivating learners to practice the second language, and including error corrections. However, Jia (2004) and Chantarotwong (2005) doubted the ability of chatbots to assist in language learning. They found that chatbots have limitations in terms of lack of memory and context awareness, expected responses, and lifeless character. However, human conversations can have the same attributes at certain times (Paltridge, 2007). Nevertheless, when it comes to pedagogy, research shows that chatbots Despite significant advancements over earlier conversational systems, chatbots appear to have much progress to make before accomplishing a goal in terms of pedagogy (Coniam, 2008a, 2008b, 2014; Williams and Compernolle, 2009), and more research is needed to explore this potential before they can be used effectively in SLA (Fryer and Carpenter, 2006).

Many studies have explored chatbots from a user interface and utility perspective (Coniam, 2008a), grammar and spelling mistakes from students' input (Coniam, 2008a), the impact of chatbots on the affective filter ((Wang, 2008), learning design of chatbots (Wang & Petrina (2013), and the impacts on student engagement (Fryer et al., 2017).

Freyer et al. (2017) investigated learner motivation. The participants were divided into two groups: experimental and control. The experimental group increasingly lost interest in interacting with the chatbot after 12 weeks of use. The explanation for this was the novelty impact and unauthentic dialogue offered by the chatbot. This conclusion contradicts previous assumptions that chatbots can inspire language learners (Fryer & Carpenter, 2006; Hill et al., 2015; Weizenbaum, 1966).

However, from an interactionist perspective, language input is not always adequate for language learners to negotiate meaning in this way. According to the interactionist theory in SAL, language learners need the opportunity to change their output when negotiating meaning (Long, 1983a; Varonis & Gass, 1985; Kramsch, 1986; Gass & Varonis, 1994). Only a few studies have examined communication between people in terms of changing output (Williams & Compernolle, 2009) or the number of negotiations for meaning in a case study (Yin & Satar, 2020). It was discovered that the chatbot's speech was void of register and had arbitrary mashups of formal and colloquial language. Additionally, the chatbot was unable to respond to multiple instances of meaning negotiation and provide sufficient remedial feedback. Williams and Compernolle (2009) contend that despite this, chatbots may nevertheless assist users in increasing their language awareness through the use of properly thought-out post-interaction activities based on certain linguistic forms identified in conversation logs, such as conversations regarding the manner in which chatbots pose questions.

Many studies have explored chatbots from a user interface and utility perspective (Coniam, 2008a), grammar and spelling mistakes from students' input (Coniam, 2008a), the impact of chatbots on the affective filter (Wang, 2008), learning design of chatbots (Wang & Petrina (2013), and the impacts on student engagement (Fryer et al., 2017).

Freyer et al. (2017) investigated learner motivation. The participants were divided into two groups: experimental and control. The experimental group increasingly lost interest in interacting with the chatbot after 12 weeks of use. The explanation for this was the novelty impact and unauthentic dialogue offered by the chatbot. This conclusion contradicts previous assumptions that chatbots can inspire language learners (Weizenbaum, 1966; Fryer & Carpenter, 2006; Hill, Ford, & Farreras, 2015).

Most research on chatbots and language learning has focused on whether the output from chatbots is linguistically correct. However, from an interactionist perspective, language input is not always adequate for language learners to negotiate meaning in this way. According to Interactionist theory in SAL, language learners need the opportunity to change their output when negotiating for meaning (Long, 1983a; Varonis and Gass, 1985; Kramsch, 1986; Gass and Varonis, 1994). Only a few studies have examined communication between people in terms of changing output (Williams & Compernolle, 2009) or the number of negotiations for meaning in a case study (Yin & Satar, 2020). It was discovered that the chatbot's speech was void of register and had arbitrary mashups of formal and colloquial language. Additionally, the chatbot was unable to respond to multiple instances of meaning negotiation and provide sufficient remedial feedback. Williams and Compernolle (2009) contend that despite this, chatbots may nevertheless assist users in increasing their language awareness through the use of properly thought-out post-interaction activities based on certain linguistic forms identified in conversation logs, such as conversations regarding the manner in which chatbots pose questions.

The literature enumerated chatbots benefits such as enhancing learners' speaking, listening skills through interacting with chatbots (Canbek and Mutlu 2016, It offers second language learners through

ASR system to improve their speaking skills. Istrate (2019) sees chatbots as virtual peers that can provide information and feedback and can help assist teachers in the classroom to enhance students' learning of speaking and vocabulary. Barcomb et al 2017 pinpoints that when there's appropriate instructional guidance from teachers, learners can be provided with a tool that can be used at any time to improve their speaking and pronunciation. They can also enhance self-learning (autonomous learning (Dizon & Tang, 2019). Despite the fact that students' perceptions in the literature are positive, their opinions regarding how chatbots understand them are contradictory. Research reports on students' satisfaction when communicating with Alexa (3.55/5) show that, compared with human raters' accuracy rate, Alexa still has difficulty understanding students' accented speech (Cardoso, 2016). This can be attributed to the way L2 students construct their questions or responses in L2, which results in a heavy cognitive load and struggling in human-chatbot interaction as a result of students struggling to access and select the relevant lexical item that is relevant to the conversation (Wu et al., 2020).

Previous literature on chatbots before the GPT era supports the fact that chatbots have the potential to improve learners' language acquisition in terms of grammar and affect, pragmalinguistic skills, and productive skills such as speaking and writing.

Studies that focused on grammar skills and affect were conducted using different types of chatbots, such as Replika (Kim, 2019), Tutor Mike (Alkhayat, 2016), Andy bot (Kim et al., 2019), and virtual tutors in game-based language learning (Wang & Johnson, 2008). Despite their inability to retain context, these chatbots were able to enhance students' language skills and reduce the affective filter, according to these studies. Due to the lack of context awareness, it may be utilized with learners with low skill levels, as it may frustrate those with great language competence.

Although there has not been much research on pragmatic strategies, a few studies have looked at the potential of chatbots and role-play to enhance pragmalinguistic skills (Oxford, 2017; Cohen, 2020). Jacquet et al. (2022) conducted a study to assess how not following conversation rules between chatbots and humans, specifically Grice's Maxims, impacted how people responded to the chatbot. The study shows that there are gender differences between male and female responses when chatbots violate Grice's maxims.

According to research, chatbots can support pragmalinguistic skills. Pragmalinguistic skills are defined as "the knowledge of the strategies for realizing speech intentions and the linguistic items used to express these intentions" (Takimoto, 2014, p. 363). Studies based on pragmalinguistic skills showed that learners were motivated and preferred interacting with chatbots. Wiske (2005) used dialogues that tested students' comprehension by providing both right and wrong answers to speech acts. After the learners finished the dialogue, they received feedback from the system. The students' ability to comprehend and employ speech acts well was demonstrated by the positive findings based on learner satisfaction and learning outcomes. Hakim et al. (2019) analyzed the strategies used by Replika to complement users in an ESL setting.

Studies have also shown that chatbots can inhibit the affective filter and enhance speaking skills. For example, Cleverbot, Let us Talk, and Talk to Eve are chatbots that were used in studies by Kim et al. (2019) and Kim (2017). These chatbots are text bots designed to handle different conversations and are more context-aware. The findings show that they reduced the affective filter and enhanced students' speaking performance.

Chatbots in the Post-LLM Era

With the advent of Open AI GPT (Generative Pre-Trained Transformer), especially GPT-3 in early 2021, the way chatbots interact with users has changed dramatically. Large Language Models (LLMs) serve as the foundation for the Generative Pre-Trained Transformer series. It was trained on billions of texts from the enormous corpus known as Common Crawl. The GPT series can produce language that appears human, respond to questions, write lengthy answers, and help with code. The parameters that govern the behavior or output of generative transformers are dependent on them. The OpenAI Playground displays these characteristics, along with others, such as temperature and presence penalty (Ray, 2023).

The GPT series has revolutionized chatbots as conversational agents and potential virtual tutors or peers for language learners. As mentioned earlier, chatbots before the GPT era were characterized by a lack of context awareness and memory. The GPT series can remember context and previous conversations thanks to the attention model. Another limitation is that AI chatbots before this era had issues in terms of linguistic usability (Coniam, 2014), persona and empathy. The Open AI Playground has enabled educators and researchers to adjust creativity, persona, and tone, making learning more personalized. (Jeon et al., 2023).

Owing to the nascent nature of the GPT series, there is still little research on its use for language learning. Dantella et al. (2023) conducted a study to examine GPT-3 linguistic usability and found that GPT-3 fails to discern subtleties in human communication. This might negatively impact students communication with chatbots without teacher supervision. Euan Bonner, a reputable researcher in language acquisition and second language learning wrote a paper with collaborators (2023) on practical usage of LLMs in the English language classroom. In their paper, they provide guidelines for language teachers on how LLMs can provide innovative ways to engage students in the classroom.

Mobile-assisted Language Learning and speaking anxiety

The ubiquity, accessibility, and affordances of smartphones have encouraged researchers to explore the potential of mobile apps in language learning classrooms. Research on MALL has shown students positive perceptions (Gamble, 2018) and interest in continuous use (Hsieh et al., 2017). Research has also shown that integrating MALL in the language learning classroom enhances vocabulary acquisition (Supti, 2019) and engages and motivates students (Khan & Islam, 2019). MALL can contribute to offering authentic materials (Octavia et al., 2019) and flexible learning (Ali et al., 2019). Systematic reviews focusing on the benefits of MALL in improving speaking skills and reducing anxiety have enumerated its contribution to lowering speaking anxiety and enhancing speaking performance (Rajendran & Yunus, 2021). For example, studies have shown that MALL can support collaborative speaking tasks (Moghaddas & Bashirnezhad, 2016; Darmi & Albion, 2017; Kusmaryani et al., 2019; Shamsi et al., 2019) and provide contextualized learning where students get the opportunity to practice in personalized meaningful contexts (Kusmaryani et al., 2019; Almarshadi et al., 2019). MALL also offers a stress-free environment that reduces performance pressure (Abugohar et al., 2017; Abugohar et al., 2019; Shamsi et al., 2019; Sun et al., 2019). It also reduces the authoritative presence of the teacher, who becomes a facilitator (Azlan et al., 2019). However, challenges remain in the use of MALL for speaking skills. Most studies have a limited scope and focus on general English contexts. Therefore, the findings of these studies may not apply to other specialized contexts in language learning. Studies have also paid less attention to the role of pedagogy and how teachers design and implement MALL-based speaking activities (Rajendran & Yunus, 2021). When it comes to ESP, there are many gaps in research that relate to the main focus of using technology in reading and writing with less focus on speaking. Research has also shown that the use of technology to improve speaking skills in the ESP context remains underutilized (Feak and Chan, 2025).

AI Agents within XR Environments

The integration of XR learning environments in the ESL classroom is revolutionizing many aspects of language learning education, as it offers immersive experiences that make impossible experiences possible for learners (EDUCAUSE, 2021). Within these immersive environments, AI promotes these experiences by creating an intelligent task-based learning space that mimics real-life language learning situations for students to practice language learning (Chiu et al., 2023).

Virtual Reality and Chatbots

While chatbot-infused VR apps are still relatively new, a few studies have examined the use of Mondly, an app for language learning that employs contextual learning to teach students in various contexts. Tai and Chen (2021) looked at how students' listening comprehension was affected by presence. They employed a combination of measures to assess students' listening comprehension, including a presence questionnaire and semi-structured interview. The study findings revealed that engaging with virtual characters, first-person perspective, and high-quality representations improved language acquisition in the VR group in terms of vocabulary retention and recalling main ideas compared to the control group that merely watched videos and failed to recall main ideas but remembered details instead.

Kawasumi and Ishii (2023) conducted a comparison study to assess how students utilized and interpreted the benefits of virtual reality vs mobile applications, with a focus on the psychological effects on their emotions. According to the study, virtual reality (VR) can enhance student motivation while simultaneously making them tired owing to the motion sickness.

Mixed Reality

Mixed reality is an immersive technology that blends the digital and real worlds (Hein et al., 2021). Despite the fact that Mixed and Augmented reality share similarities in terms of blending the digital and real worlds, they differ in terms of immersive interactions. MR provides more interactions with and manipulation of digital objects than AR (Panagiotidis, 2021).

In the past, Mixed Reality (MR) was too costly for research use, as MR devices such as Hololens and Magic Leap were very expensive. However, the emergence of competitive headsets, such as Pico 4, which challenged Meta, led to the release of Meta Oculus 3 and a new Meta SDK. This development has simplified the process for developers, suggesting that MR will become mainstream, with more MR apps anticipated to launch.

Research on MR and language learning provides some insights into the advantages of MR., with the literature frequently mentioning research conducted using Hololens 2 in the context of language learning. Leonard and Fitzgerald (2018) conducted a pilot study in a secondary school in Australia and found that MR resulted in effective interactions among students. In their paper, they pointed out the advantages of MR over VR in learning design. They highlighted that MR technologies “offer mobility, and therefore sensory– motor engagement in the real world” (p.3).

Vazquez et al (2017) conducted a study that focused on learning vocabulary using Microsoft HoloLens. They reported the advantage of MR as it blends holograms with the physical world to provide more contextual learning opportunities than VR. This study suggests a framework based on serendipitous learning in mixed reality to improve learners' language acquisition skills.

Learning Affordances of Mixed Reality

Immersive Learning has become increasingly popular in different fields, including medical training, second language teaching, and STEAM education (Bower et al., 2017; Chang et al., 2023; Maas & Hughes, 2020) as it offers a safe immersive environment where learners can have the freedom to fail and repeat the experience until they master learning. MR allows users to interact with the digital and physical worlds without being isolated from classroom tasks.

When it comes to integrating conversational AI into XR technologies, research shows it to be a promising tool that can tackle challenges in education. However, the literature has mainly focused on conversational AI agents in VR (Liaw et al., 2023; Dang et al., 2025).

Challenges and Research Gaps in ESP

ESP focuses on the professional aspects of the language, rather than general English courses that focus on daily life expressions (Fitria, 2019; Paltridge & Starfield, 2014). ESP's main challenges are that most research is overly focused on writing and reading skills rather than oral performance, which makes it challenging to design speaking curricula (Paltridge & Starfield, 2013).

What contributed more to this imbalance can be attributed to three causes: students, teachers, and environmental causes (Ghafar, 2022). As for teachers, these limitations include insufficient specialized teacher training (Ho, 2011), lack of authentic teaching materials (Lam, 2011), ineffective teaching strategies (Ghafar, 2022; Ho, 2011), lack of theoretical foundations to support ESP instruction (Chen, 2011), and focus on one professional domain that relates to teacher knowledge.

Regarding challenges related to ESP students, research enumerates several challenges (Ghafar, 2022). These challenges can be related to students' inadequate exposure to professional vocabulary that is rarely used in real-life situations. Heavy focus on grammar and overreliance on dictionaries lead students to have limited language use and underdeveloped skills, especially in terms of speaking (Rezaei et al., 2012). As for environmental causes, research shows that the exam-oriented nature of ESP courses has emphasized memorization over practical applications. Lack of instructional resources (Paltridge & Starfield, 2013), limited skill practice, and inadequate materials have contributed to students' less engagement and reduced motivation (Ghafar, 2022). Therefore, addressing these challenges and limitations requires clear theoretical and sound pedagogical approaches that can serve as a framework for technological intervention using LLMs to potentially complement teacher knowledge and provide learners with personalized, situated instruction.

.

Theoretical Framework

The Interactionist Hypothesis and Negotiation of Meaning

The interactionist hypothesis is an SLA theory influenced by the constructivist school of thought (Sarem & Shorzadi, 2014). The constructivist perspective transcends Chomsky's innateness hypothesis and the cognitive psychological perspectives in stressing the importance of individuals constructing their reality and engaging in "in social practices…on a collaborative group or on a global community" (Spivey,1997, p. 23, cited in Brown, 2000). In the literature, Vygotsky and Piaget have different opinions on the role of social interaction in language acquisition. Piaget explained that improvement occurs in a learner's brain in predetermined stages, where social interactions are triggered when the brain is ready to learn (Yildiz, 2025). In contrast, Vygotsky emphasized the importance of social interactions as fundamental to cognitive development, rejecting the notion of predetermined stages. For Vygptsky, social interactions construct meaning and occur when the learner interacts with their environment (Yilidiz, 2025). Based on Vygotsky's view that social interaction enhances learning, the interactionist hypothesis (Long, 1996) was developed based on constructivist theory, underscoring that negotiation of meaning facilitates learning through communicative exchanges.

According to the interaction theory, the term interaction refers to the learner's interaction with peers, teachers, or native speakers. In this process of interaction, for the interlocutors to arrive at mutual understanding, this can become possible through negotiation of meaning, where they ask questions for clarification, repetition, and make sure that they understand each other. Long defines negotiation of meaning as follows:

> The process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, message content, or all three, until an acceptable level of understanding is achieved.
>
> (1996, p. 418)

According to Long, the negotiation of meaning is important for language acquisition for several reasons. It involved competent speakers who interacted with language learners who modified their conversations to make them comprehensible to ESL learners. Negotiation of meaning involves input, what the learner hears; selective attention, what the learner pays attention to; and output, what the learner says, to improve language acquisition. It contains relevant linguistic input and semantically connected discourses. Negotiation of meaning involves input modifications, such as clarifications and repetitions, that contribute to clarity and context.

Negotiation of Meaning in Synchronous CMC

NoM in text chat SCMC has been widely researched since the 1990s (see, for example, Chun, 1994; Ortega, 1997; Blake, 2000; Pellettieri, 2000; Fernandez-Garcia & Martinez-Arbelaiz, 2002; Smith, 2003, 2004; Tudini, 2003; Akayoglu & Altun, 2009; Samani et al.,, 2015) as learners can edit and modify their output while interacting by sending messages to others. In the absence of gestures and facial expressions, the entire meaning is conveyed by the written text. Pica (1994) divided negotiation of meaning strategies into three types: clarification requests, confirmation checks, and comprehension checks. Confirmation checks are performed to ensure that the speaker's statements are understood by the listener. Comprehension checks occur when the learner summarizes or repeats what they hear. Clarification requests occur when a learner asks the speaker to repeat or clarify a confusing statement.

The efficiency of NoM depends on several factors, including the learner's proficiency level, task complexity, and learning context. Research that examined learner-voice-based NoM strategies (Kim, 2017) in comparison to student-student NoM strategies found that students' proficiency level impacted the type of NoM strategies the students used. Moreover, students in the chatbot condition used more strategies than those in the traditional condition. The increase in NoM strategies in the chatbot group allowed for improvement in students' performance. Yin and Satar (2020) also explored NoM routines with students that had different proficiency levels using two different chatbots: a pedagogical chatbot, and a conversational chatbot. They discovered that students with high proficiency levels expressed more interest and produced more NoM patterns when using conversational rather than pedagogical chatbots.

The problem with this is that the study was conducted in 2020, which was the pre-LLM era, in which the chatbot they selected was scripted and rule-based and could have decontextualized or mis-contextualized responses. Conversational chatbots were designed using Artificial Intelligence Markup Language that uses pattern NoM strategies to help understand how multimodal interaction with CAI agents improves learners' speaking performance and how this interaction contributes to learners' sense of presence in an immersive environment. Research on CMC shows that negotiation strategies with conversational agents help improve lexical (i.e., vocabulary) development and fluency (Lee, 2020; Sauro, 2011). In immersive or AI-enhanced contexts, multimodal cues such as gaze, tone, and gestures can further scaffold these negotiation processes or trigger more frequent negotiation strategies than the voice-based chatbot (Lan, 2020; Yamada & Akahori, 2007)..

Interaction with AI Agents and Speaking Performance

Incorporating conversational AI, whether embodied or voice-based, within the interactionist framework provides opportunities to explain how AI agents can enhance oral performance through interaction where the NoM takes place. When learners engage in dialogue with conversational AI agents, especially those with human-like features such as gestures, gaze, voice, and facial expressions, they experience a sense of communicative authenticity, which prompts them to adjust their discourse strategies, improving syntactic complexity, fluency, and vocabulary (Lan, 2020; Lee, 2020). This interaction process allows for the negotiation of meaning in which AI agents provide relevant feedback, answer clarification requests, and confirm checks that replicate the scaffolding techniques offered by teachers but in a safe, anxiety-free environment. Accordingly, it sustains interactional flow and promotes modified output in which learners realize the errors they made and work on resolving communication breakdowns (Long,1996; Shikun et al., 2024). Moreover, the embodied nature of the AI agent can enhance social-cognitive interaction by simulating the dynamics of real conversations, which can encourage students to experiment with language without fear and lower affective barriers (Moreno & Mayer, 2007; Dewaele & Dewaele, 2020). Research has reported an increase in performance (Koç & Savaş, 2024; Yin & Satar, 2020) in students interacting with AI agents (22%) (Naseer et al., 2024).

However, many studies have focused on text-based or voice-based chatbots, leaving multimodal ECAs underexplored in relation to NoM, speaking performance, and anxiety.

In sum, NoM strategies, according to the interactionist hypothesis, drive acquisition as they create opportunities for modified input, making input comprehensible, helping co-construct meaning, and promoting cognitive development. These processes align with pedagogical approaches such as Task-based Language Teaching (TBLT), which are learner-centered, employ authentic approaches, and emphasize meaningful language use rather than drills and mechanical practice. TBLT promotes scaffolding, interaction, and NoM within purposeful tasks.

Task-Based Language Teaching (TBLT)

Although it is not a learning theory, TBLT is a teaching approach that is primarily based on learning theories rather than on linguistic theories about grammar and structure. However, it considers language as a medium in which tasks are initiated, meaning is negotiated, and learning outcomes are achieved. Thomas and Reinders (2010) observed that TBLT and CALL share common grounds when it comes to learning methods such as “project-based, content-based, and experiential learning”, and theoretical approaches, such as “constructivist and social constructivist thought” (p. 5).Task-based Language Teaching is a teaching method rooted in Communicative Language Teaching (CLT), which emerged as a response to traditional teaching approaches (Ellis, 2003; Long, 2015). TBLT emphasizes the role of activities that focus on authentic, goal-oriented tasks that mimic real-world contexts and incorporate meaningful engagement employing communicative tasks rather than isolated drills (Nunan, 2004). Through these activities, learners interact, negotiate meaning, produce comprehensible input, and obtain feedback (Long 1996).

TBLT is a teaching method that is divided into three sequential stages. The pre-task phase focuses on learner schema to prepare learners cognitively and linguistically for the main task (Van den Branden 2006). The teacher's role is to provide clear instructions, link learners' prior knowledge to the existing lesson, and offer preparation activities such as warmup activities (Ellis, 2004). At this stage, the focus is more on vocabulary than on grammar through guided activities. Research shows that this stage helps improve language complexity, but has uncertain effects on accuracy (Philip, Oliver, & Mackey, 2006). The second stage is called the main task or performance stage, where learners become active participants or risk-takers involved in pair or group activities. At this stage, the teacher's role shifts to that of a classroom facilitator or monitor (Ellis, 2004), where they provide guidance and ensure that the activity is learner-centered. Task types in this phase vary widely, including information gap tasks and jigsaw tasks, which require negotiation and interaction to reach solutions. The final stage is the post-task stage, where learners review what they learned through three main methods: reporting, repeating, or evaluating. To explain, learners review their understanding by reporting what they accomplished through the main task (Johnston, 2005, as cited in Willis & Willis, 2007), repeating the same task but under stricter time limits (Ellis, 2003; Pinter, 2005; Bygate, 2001; and Essig, 2005; all cited in Willis & Willis, 2007), or reviewing the gaps in their language use through text recycling and self-correction (Motlagh et al., 2014).

In Technology-mediated and immersive environments, TBLT can be a strong framework that integrates immersive environments incorporating ECAs. Research exploring the role of immersive environments in language learning employing TBLT has shown that VR-based tasks (e.g., making an appointment or asking for directions) enhance learner engagement and interaction compared to less authentic tasks in the traditional classroom (Christoforou et al., 2024). Virtual environments allow for situated learning that bridges the gap between traditional classroom instruction and real-life language usage. Hence, it aligns with TBLT, which creates complex, authentic tasks where learners can negotiate meaning, resolve communication breakdowns in the target language, and improve their performance (González-Lloret & Ortega, 2014).

When immersive environments are combined with embodied multimodal conversational agents within the TBLT framework, ECAs serve as interlocutors or task members that interact in scenarios that mimic real-life workplace or social interactions. Interactive tasks in immersive learning environments have shown improvements in speaking (ElShazly, 2021; Ogawa, 2025; Yan et al., 2024). However, the alignment of task design with language outcomes in immersive environments remains unexplored. Moreover, evidence on AI-powered immersive environments impact on anxiety remains inconclusive (Wiboolyasarin et al., 2025)

The importance of TBLT in this study can inform both task design and how ECAs and immersive environments can help scaffold learners' negotiation of meaning, output, and feedback cycles. Immersive environments transform interaction from the plane to three-dimensional and embodied communication, where learners utilize multimodal resources and verbal language to negotiate meaning, which is essential for TBLT (Chen & Sevilla-Pavón, 2023; Smith & McCurrach, 2024). By situating language learning in context-rich, multimodal, and goal-oriented tasks, this framework supports the examination of how AI-mediated interactions can enhance speaking performance and influence learners' anxiety and sense of presence. Thus, the theoretical alignment between TBLT and the interactionist hypothesis guided the design, implementation, and analysis of the learning tasks in this study (Ellis, 2003; Long, 1996; González-Lloret & Ortega, 2014; Chen & Sevilla-Pavón, 2023). Situated learning theory complements this perspective by positioning learning as intertwined with the social and contextual environments in which communicative activities occur (Yao et al., 2024).

Situated Learning Theory

Situated Learning Theory posits that learning is best achieved when it occurs in the same context and social communities in which the learner participates (Lave & Wenger, 1991). Traditional classrooms can be challenging in terms of treating learning as a form of decontextualized activity (Brown et al., 1989). According to situated learning theory, the learner, task, and environment are elements that contribute to shaping and impacting each other (Lave & Wenger, 1991).

Regarding SLA, situated learning theory offers a valuable theoretical framework for understanding how learners develop communicative competence by interacting with conversational agents in immersive environments. In addition, limited peripheral participation allows learners to gradually move from an observer role to active engagement (Lave and Wenger (1991), which aligns with the goals of TBLT, integrating meaningful instruction within authentic, context-based activities that mimic real-world experiences rather than activities that prioritize memorization or drilling practice (Warschauer et al., 2000). This, in turn, encourages learners to internalize linguistic structures and helps them become active constructors of knowledge (Felix, 2002).

In immersive learning environments, SLT offers a theoretical lens for understanding how technology can simulate authentic communicative settings. By recreating immersive environments that simulate realistic scenarios, learners can engage with language in situated, socially meaningful ways (Yanet al., 2024). Incorporating ECAs provides learners with the opportunity to interact with interactive partners that enable legitimate peripheral participation. In this learning environment, learners practice language within a judgement-free social context, where ECAs serve as mentors, scaffolding learning, and drawing knowledge from a large body of text that helps complement teacher knowledge while maintaining the authenticity of communicative practice (McGivney, 2025). Thus, learning through participation, meaningful interaction, and authentic communication exemplifies the principles of SLT and TBLT.

When learners are required to navigate linguistic challenges in an environment that fosters a situated context and meaningful interaction with ECAs in well-designed meaningful tasks, this can foster their language development (Nunan, 1989; Shih & Yang, 2008). However, research on the application of SLT and TBLT in immersive environments remains underdeveloped (Özüdogru & Özüdogru, 2017; Yan et al., 2024). Thus, a situated approach to AI-enhanced immersive language learning environments based on authentic, task-based participation offers a promising direction for understanding how these types of environments can help learners improve their speaking performance and build communicative competence.

Social Presence Theory

SP theory explains how humans interact with machines that have human-like characteristics through conversations that influence the “sense of being with another” (Biocca et al., 2003, p. 456), where they feel connected to a communication partner (Oh, 2018). Studies on human-computer interaction demonstrate that a more richly developed medium improves the feeling of social presence (Daft et al., 1986; Hassanein and Head, 2007) compared to less sophisticated media (Lee et al., 2009). In educational settings, social presence enhances engagement and meaningful interactions, which are important for effective learning (Richardson & Swan, 2003; Lomicka & Lord, 2007).

Studies exploring social presence in multimodal communications have underscored the importance of gaze in promoting social presence (Satar, 2013). Research exploring social presence in terms of technology affordances, multimodal features, and student traits found that students from different domains had more positive perceptions of SP when interacting through technologies such as video conferencing (Yamada & Akahori, 2007), with avatars in virtual environments (Wang et al.,2016), pedagogical agents through video compared to text (Borup et al., 2014), or control groups (Huang et al.,2021). Furthermore, research has found that multimodality can have positive effects on SP. For example, Want et al. (2019) demonstrated that teachers' guidance to students through gaze in a video increased SP and impacted students' learning. When it comes to avatars or embodied conversational agents in immersive environments, research shows that photorealistic and anthropomorphic avatars that mimic human behavior through gaze, voice, and gestures result in stronger SP (Oh et al., 2018).

. However, research cautions that designing an AI-enhanced immersive environment where AI interactions do not align with the learning goals, context, and learners' expectations can lead to negative effects (Harris et al., 2009, 2023).Therefore, more research on AI-enhanced immersive environments and SP is needed to understand how to create meaningful social experiences (Oh et al., 2018).

Foreign Language Anxiety

The following section addresses the theoretical lens that incorporates both subjective and objective frameworks to examine FLA. First, it outlines how self-reported measures capture learners' subjective experiences of anxiety during communicative tasks, particularly in immersive and AI-mediated environments such as VR. Second, it introduces physiological measures, such as heart rate monitoring, as an objective means of assessing emotional arousal associated with language learning tasks. Together, these complementary perspectives provide a comprehensive foundation for understanding the affective dimensions of language learning in immersive environments.

Horwitz FLA theory

When interaction takes place, anxiety in foreign language learning can inhibit learning. Horwitz et al. (1988) suggested that anxiety in the foreign language classroom (FLA) is due to three reasons: a) fear of negative judgement by others, b) fear about communication, c) test anxiety which stems from the fear to fail. The Horwitz FLA anxiety survey has been used as a framework to assess subjects' anxiety in pre- and posttests, but several studies in education have criticized the use of self-reports, especially with non-native speakers in "multiethnic settings, which can be difficult due to researchers lack of understanding of respondents' cultural norms and misunderstanding of survey questions. while heart rate measures can offer a more objective measurement of Autonomic Nervous System (ANS) function that would result in reduced measurement error and real-time monitoring, which is less affected by cultural differences in terms of communication (Hunt & Bhopal, 2004)..

Horwitz et al. (1986) claimed that FLA is mainly related to three performance related factors: (a) fear of negative evaluation or apprehension about others' evaluation, (b) communication apprehension, and (c) test anxiety (i.e., performance anxiety caused by fear of failure).

This study aimed to understand the interplay between self-reported measures of anxiety and heart rate (HR) as indicators of cognitive performance in the second-language learning classroom. The theoretical framework guiding this study is grounded in the biopsychosocial model, which posits that anxiety can impact cognitive outcomes by interacting with physiological responses (Engel, 1977). Research has shown that anxiety disorders are associated with physiological markers of stress (Laborde, Mosley, & Thayer, 2017). The aim of integrating self-reports and HR is to comprehensively understand how participants' subjective experiences relate to physiological responses and, ultimately, cognitive performance. This dual approach not only enhances the validity of this study but also aligns with researchers' recommendations for using multiple data sources to capture the complexity of anxiety and its effects on learning (Kwon et al., 2017). While self-reports provide insights into learners' subjective emotional states, research shows that task design and interaction with peers or AI-mediated interlocutors can influence reported anxiety levels, with some students reporting high anxiety rates during complex tasks or due to cognitive load while using novel tools (ElShazly, 2021; Ogawa, 2025; Yan et al., 2024). Therefore, it is crucial to incorporate an objective lens to fully understand learners' emotions while performing communicative tasks.

Physiological Arousal in Language Anxiety

Heart rate (HR) provides objective measures that complement self-reported data. HR has been used to index arousal and emotional responses in immersive learning and simulation studies (Edelmann & Baker, 2002; Gorini et al., 2010; Slater et al., 2006). HR can record an aspect of learner experience that might not be captured by self-reports, allowing for triangulation between subjective and objective indicators of anxiety.

Research shows individual differences in emotional and physiological reactivity among learners during demanding communicative tasks, where some exhibit high anxiety while others show no consistent correlation with self-reports (Gorini et al., 2011; Slater, 2004; Yan et al., 2024).Therefore, there is a need to combine self-report instruments with physiological measures to obtain a more comprehensive understanding of learners' affective states during AI-enhanced immersive-based tasks.

This study adopted a dual approach to address how anxiety impedes communicative competence in speaking performance. Horwitz et al. 's (1986) Foreign Language Classroom Anxiety Theory was used to understand situation-specific anxiety using the Foreign Language Classroom Anxiety Scale (FLCAS) (Horwitz, 1986), which has been extensively studied and widely applied (Mak, 2011). However, reliance on self-reports raises concerns, especially in multiethnic settings, where cultural norms or question misinterpretations can skew the findings (Sudina, 2021). Research has mentioned the limitations of self-reports in terms of lacking the ability to measure anxiety in real time (Shachter et al., 2022), recall bias (Solhan et al., 2009), and social desirability bias (Latkin et al. 2017, Ryan et al., 2021). Research also shows that during stressful activities such as public speaking or speaking a new language, the sympathetic nervous system is activated and the heart rate increases. When anxiety decreases, the parasympathetic system reduces the heart rate. Due to such automatic responses from the autonomic nervous system, heart rate measurements can offer the potential to examine what learners feel in real time (Gregerson et al., 2014). Therefore, triangulating subjective and objective measures of anxiety can contribute to an accurate understanding of students' feelings before and after the intervention. This can also contribute to understanding how students can perform well in speaking performances and what causes them to struggle (Ganster et al., 2018; Hardy, 1999; Schmidt et al., 2020; Teigen, 1994).

Chapter III

Methodology

Introduction and Overview

This quasi-experimental, mixed-method design study investigated and compared two conversational AI agents as supplementary tools in the classroom to reduce students' anxiety and improve their speaking performance. The study also examined the way students negotiate meaning when interacting with AI for six weeks to see how their questions improve their understanding of job interview questions assigned to them. Moreover, the qualitative component of the study in which the students sit for reflection interviews shows their perceptions about each condition (i.e., control, ChatGPT voice app, conversational AI agent in MR) in terms of how it reduced their anxiety, improved their performance, and their perceptions about classroom teaching in general and how the integration of AI helped them.

The research questions examined were as follows:

1- How does interacting with an AI agent in Mixed Reality, the ChatGPT voice app, or traditional methods impact participants' overall speaking performance across seven rubric dimensions (fluency, pronunciation, intonation and stress, grammar and sentence structure, vocabulary use, content and relevance of answers, and use of work-based examples) from pre- to post-mock job interviews, and which interaction modality yields the best overall speaking performance?

2- How do interactions with an MECAI agent in mixed reality, the ChatGPT voice app, and traditional methods impact participants' self-reported anxiety levels and heart rates during the pre- and posttest mock job interviews conducted by an external examiner?

3- What are the different NoM strategies used by students to negotiate meaning when interacting with AI agents in mixed reality or using the ChatGPT voice application?

4- What are the perceptions of the students in the experimental and control groups about the effectiveness of these conditions in helping them overcome their anxiety and improve their speaking performance?

This chapter describes the study design, sample, and population. This chapter also explains the data collection procedures and rationale for the methodology and design. Instruments used, data analysis, and literature underpinning the selected methods. This study employed a quasi-experimental mixed method design approach to triangulate multiple data collection sources to examine the effect of the intervention on the two experimental groups compared to the control group and to determine whether there were significant differences in speaking performance as well as anxiety reduction. Three pre- and posttest mock job interviews, anxiety self-report (FLCAS), and heart rate measurements assessed 55 participants' speaking performance and anxiety when interviewed for a hypothetical mock job interview by an external examiner. The study also examined students' interactions with both treatments (ChatGPT voice app and conversational AI agent in MR) during the six-week intervention by examining the negotiation of meaning, in which students used strategies such as confirmation checks, repetitions, and clarification requests to improve their speaking skills.

As for the qualitative aspect of the study, the study also examined perceptions of 36 participants from all three groups to reflect on the whole experience in general, as they were asked 11 questions to examine their experience within their assigned condition and to reflect on challenges they had in terms of practicing job interview questions in a second language class teachers' role, assigned condition role in reducing anxiety and improving speaking performance, and how this impacted them academically and emotionally. A mixed method design was used to provide a complete picture of how the treatments affected the participants' performance and anxiety during mock job interview tasks.

Research Sample

This quasi-experimental mixed-method design study will be conducted from January 2025 to November 2025. The study sample consisted of 55 ESL students aged between 17 and 64 years old attending the Skills for Success 2 class at Hudson County Community College (HCCC). The participants in this study were all at the A2 (Elementary) level of English proficiency. Their assigned levels were based on their placement test scores in listening (41-61) and grammar (41-60), which aligned with established proficiency scales (e.g., TOEFL iBT 41-65; IELTS 5.0-5.5; CEFR 41-65). Most students have been introduced to job interviews at level A1 without being introduced to the mock job interview activity that is required at level 2. The study was conducted at HCCC across two semesters, Spring and Fall. HCCC encompasses a wide range of multicultural and socioeconomic backgrounds in New Jersey, including both North Hudson and Jersey City HCCC campuses. The study was conducted in the researchers' classes as well as another instructor's class after obtaining approval from the Institutional Review Board. Students in the researcher's class took the pre- and posttest mock job interviews with an external examiner who signed an informed consent form, whereas the students who participated in the study from another classroom run by a different instructor were interviewed for the pre- and posttest mock job interviews by the researcher herself as their external examiner. All students, participating external examiners, and classroom instructors signed informed consent forms. Table 1 provides a statistical summary of the sample population, covering key demographics as well as participants' backgrounds in English, job application history, and interview preparedness.

Table 2

Study Demographic Questions Responses

Demographic Questions Responses		
Age (years)	**Mean (SD)**	29.20 (13.65)
	Median (IQR)	23.00 (19.00, 40.00)
	Min-Max	17.00 – 63.00
Gender	**Females n (%)**	42 (76.36)
	Males n (%)	13 (23.64)
Duration of residence in the USA (months)	**Mean (SD)**	37.13 (46.23)
	Median (IQR)	24.00 (8.00, 42.00)
	Min-Max	2.00 – 240.00
Speaking English at home	**Yes n (%)**	7 (12.73)
	No n (%)	48 (87.27)
Country of origin	**Afghanistan n (%)**	1 (1.82)
	Algeria n (%)	2 (3.64)
	Argentina n (%)	1 (1.82)
	Colombia n (%)	3 (5.45)
	Dominican Republic n (%)	15 (27.28)
	Ecuador n (%)	12 (21.82)
	Egypt n (%)	7 (12.73)
	India n (%)	1 (1.82)

	Mexico n (%)	1 (1.82)
	Morocco n (%)	2 (3.64)
	Peru n (%)	6 (10.91)
	Senegal n (%)	2 (3.64)
	Sudan n (%)	1 (1.82)
	United States n (%)	1 (1.82)
Previously applied for job in the US	**Yes n (%)**	41 (74.55)
	No n (%)	14 (25.45)
Having confidence speaking English	**Very confident n (%)**	8 (10.91)
	Somewhat confident n (%)	17 (30.91)
	Neutral n (%)	21 (38.18)
	Somewhat unconfident n (%)	6 (10.91)
	Very unconfident n (%)	3 (5.45)
Previous formal training on English	**Yes n (%)**	22 (40.00)
	No n (%)	33 (60.00)
Number of months of learning English	**Mean (SD)**	8.95 (13.87)
	Median (IQR)	6.00 (3.00, 8.00)
	Min-Max	1.00 – 96.00

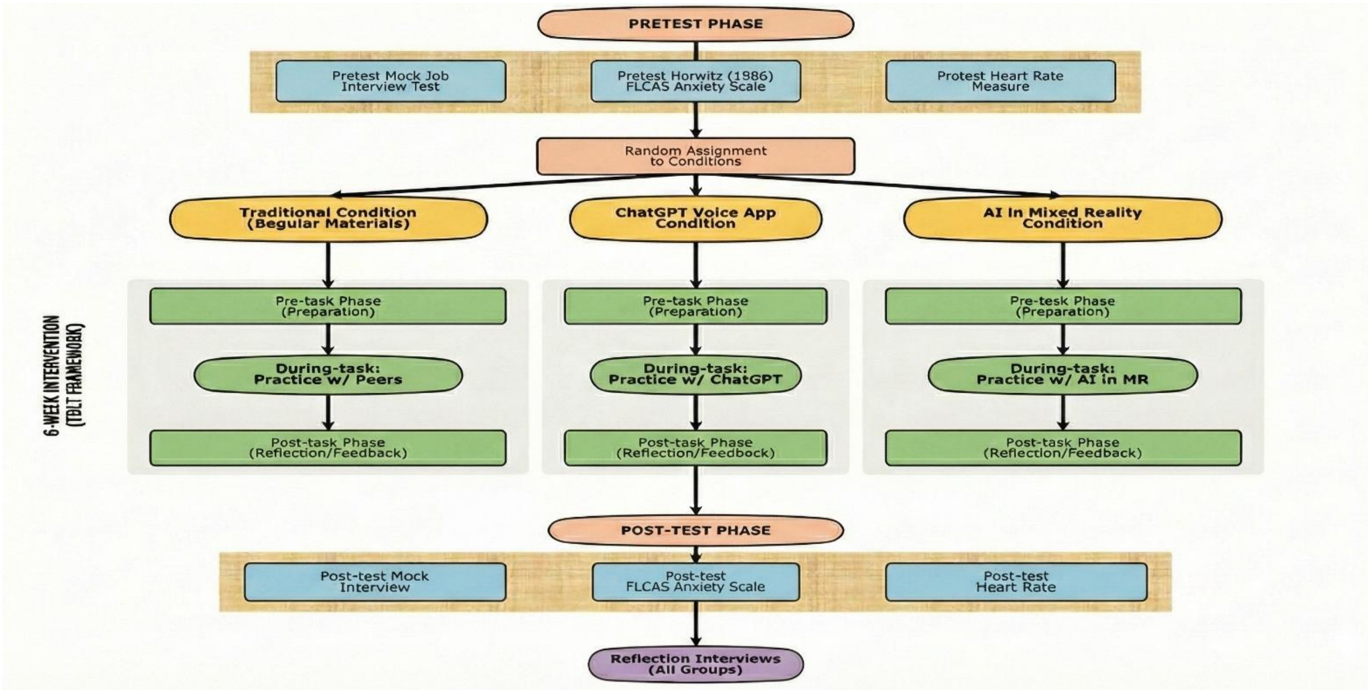
Table 3

Students' Conditions

As shown in Table 3, the students were randomly assigned to three conditions. The procedures for each condition are illustrated in Figure 1.

Condition	Traditional Group	ChatGPT Group	MECAI Group
Explanation	The group uses traditional classroom materials	The group uses ChatGPT voice application	The group interacts with an AI agent using HMD

Figure 1 *Study Demographic Questions Responses*



1. *Demographic survey and FLCAS pretest for all groups*: The same procedure was employed for each class involved in the study. After all students completed the consent forms, they were provided with the Horwitz FLCAS (1986) survey, which included demographic questions. The demographic survey included questions about gender, age, language spoken, whether English was spoken at home, whether the students felt confident during job interviews, and whether the students had formal job interview training before. All students were asked to prepare for a hypothetical job interview to prime them for the main task (i.e., the intervention). They were asked to write down their dream job title on a sheet that would be given to the external examiner and prepare for a mock job interview (i.e., the pretest).

2. *Mock job interview pretest*: The following week, the participants were asked to sit for a mock job interview with an external examiner to test their prior knowledge. During the interview, each participant wore a Polar OH1+ device to measure their heart rate. A baseline reading was taken for

two minutes before the interview began. After students completed their pretests, they were randomly assigned to one of three conditions: two experimental groups, the MECAI in MR and the ChatGPT voice app, and the control group, the traditional condition. After being assigned their conditions, the experimenter provided training to both experimental groups. The ChatGPT voice app group was trained to use the ChatGPT voice application and was given the prompt they should use to start training (Table 6). The MECAI in the MR group were shown how to use the Oculus Pro headset and controllers to control how to speak or interrupt the AI agent.

3. 20- Minute *Interaction with AI for Six Weeks*: The ChatGPT group was also trained to send their conversation logs to the experimenter via the classroom Canvas. The MECAI in the MR group was not required to send the experimenter their conversation logs as the experimenter had access to the conversation logs saved on the Convai website.

4. *Posttests*: After the six-week intervention, students completed the posttests: they all filled out the posttest Horwitz (1986) anxiety FLCAS and sat for the posttest mock job interview, during which they wore Polar OH1 to measure their heart rate.

5. *Reflection interviews*: In these interviews, students were asked about their perceptions of the conditions to which they were assigned. In a recorded interview, they responded to 11 questions about their perceptions of their assigned conditions and their impact on their speaking performance and anxiety levels.

Data Collection Measures.

Demographic survey

The researcher developed a demographic survey to collect information on the participants, including age, gender, years of studying English, language used at home, and previous formal job training. These items served descriptive purposes and were not intended to measure any constructs. Therefore, validating the demographic survey was not possible. A demographic survey was provided to students in the posttest to account for any differences that occurred after the six-week intervention.

Mock Job Interview Questions

Eight questions were adapted from Lou et al. (2010) (See Appendix A). Questions were selected and modified to cater to students' diverse professional interests. The questions included items that ranged between questions about interest in the job, weaknesses, and strengths. The questions were administered in pre- and post-mock job interviews by an external examiner before and after the intervention. To ensure fairness and avoid bias across all conditions, external interviewers signed consent forms to participate in mock job interviews as external examiners. They interviewed students in pre- and post-mock job interviews. Thus, ensuring that familiarity with classroom teachers does not intervene with anxiety levels and ensuring that all students have equal experience. The questions were then rated using a rubric adapted from Ramanarayanan et al. (2017). However, Ramanarayanan et al. 's study focused only on pronunciation, fluency, and intonation, and four more items were added to align to expected responses in a job interview, mainly grammar and sentence structure, Content & Relevance of Answers, Use of Work-Based Examples, Vocabulary Use. It was necessary to add these items to the rubric so that raters could examine how the learners would answer in the pretest and how they would apply what they learned in the posttest. Three experienced and independent raters, including the researcher, rated the pre- and posttest mock job interview questions, and interrater reliability was assessed using Fleiss' Kappa. According to the guidelines of Landis and Koch (1977), a substantial agreement of k=75.

Horwitz et al. (1986) FLCAS Self-Report

Horwitz's et al., (1986) Foreign Language Anxiety Scale is employed as a pre- and posttest prior to and following the intervention. It consists of 33 questions on a 5-point Likert scale (See Appendix C) designed to evaluate learners' foreign language learning anxiety in the classroom. The responses ranged from 1 (Strongly Disagree) to 5 (Strongly Agree), with possible scores ranging from 33 to 165. The self-report measure assesses three types of anxiety: communication apprehension (speaking anxiety), test anxiety, and fear of negative evaluation. Higher scores indicate greater anxiety levels. The FLCAS self-report has been extensively used in Applied Linguistics and language learning research. Of the 33 questions, 11 were negatively worded and reverse-coded in line with standard practice to ensure consistency (Al-Saraj, 2014; Bojović, 2024).

Heart Rate Measures

This study employed a Polar OH1+ heart monitor to measure heart rate data during the pre- and posttest mock job interview. Polar OH1+ has been validated in both sedentary and active contexts. A recent validation study showed the high accuracy and reliability of Polar OH1+ across diverse activities, including sedentary to high-intensity activities, with a concordance correlation coefficient of 1.00, which included a mean bias of 0.05 BPM (Schweizer & Gilgen-Ammann, 2025). Each participant wore a Polar OH1 + device on their forearm during the pre- and post-mock job interviews. First, participants were asked to relax for two minutes so that the researcher could measure their baseline. Following the two-minute resting period, the participants immediately started their mock job interview. Measuring the baseline period before task onset is consistent with established psychophysiological research that differentiates between baseline arousal and task-related physiological responses (Kao et al., 2014; Lehrer et al., 2007).

Conversation Log Analysis for NoM

To analyze the students' interactions with both AI agents, Computer-Mediated Discourse Analysis (CMDA) was employed (Herring, 2001). Two raters were involved in the rating procedure, following Long (1996) and Pica (1994). .Computer-Mediated Discourse Analysis (CMDA) was employed to examine research question 3 for three reasons: (a. It examines the discourse underlying patterns, whether explicitly or implicitly (b). It involves speakers' communication decisions and their choices. These communication decisions are affected by technological features. In CMDA, researchers focus on how technology shapes the way users communicate and interact, and whether technology can alter the way they express themselves and interact. As both experimental group conversation logs had different counts, the frequencies of negotiation of meaning were normalized per 1,000 words to control for the differences in total word production. Normalization was calculated using the formula shown in equation 1 below:

Equation 1

Normalization Equation

$$\text{Normalized Frequency} = \frac{\text{Observed Frequency of Negotiation of Meaning}}{\text{Total Words in Chatlog}} \times 1000$$

The CMDA focuses on interaction patterns in digital communication. It is used in research to examine how participants negotiate meaning and manage communication breakdowns during computer-mediated communication. In this study, two experimental groups, the ChatGPT and MECAI in the MR group, interacted with different conversational AI tools to improve their speaking skills within a specific activity. Accordingly, CMDA was employed to identify how learners negotiate meaning with different conversational AI agents and what negotiation patterns they use to negotiate meaning across synchronous (real-time) voice interactions and multimodal computer-mediated communication. To do so, a hybrid coding framework from Patterson and Trabaldo (2006) was expanded with four additional categories identified by Akayoglu and Altun (2009). These combined categories have helped identify key patterns of NoMs, such as clarification requests, confirmations, elaborations, and vocabulary checks. The complete set of NoM patterns with examples is provided in Tables 4 and 5.

All conversation logs were transcribed, anonymized, and coded manually. Two experienced raters were involved in the coding process of the interviews. Each exchange between the learner and AI was treated as a unit of analysis and assigned one function of negotiation of meaning. To ensure consistency, 20% of the data were calculated for interrater reliability, which resulted in substantial agreement (Cohen's K =.82). The experimenter then conducted frequency counts to explore the NoM pattern that occurred across six weeks per group.

Table 4

NoM Taxonomy and Functions adapted from Samani et al. (2015)

Taxonomy by Patterson and Trabaldo (2006)	NoM Added by Akayoglu and Altun (2009)
Clarification request	Confirmation*
Comprehension check	Elaboration request*
Confirmation request	Reply elaboration*
Correction/ self – correction	Elaboration
	Reply clarification / definition
	Reply comprehension
	Reply confirmation
	Reply vocabulary
	Vocabulary check

* NoM Functions added by Akayoglu and Altun (2009) in Samani et al. (2025)

Table 5

Functions of Negotiation of Meaning: Explanations and Examples

Function	**Explanation**	**Example**
Clarification request	The listener does not understand the speaker and asks questions such as "I don't understand".	Student: What do you mean?
Confirmation	The listener checks understanding by repeating what they heard.	Student: I said I work in hospital in my country not in US.
Confirmation check	Student asks if what they understood is correct.	AI: Can you work weekends? Student: Weekends?
Correction or self-correction	A student correcting themselves after they realized they made a mistake.	Student: I no have a license. I don't have* a license
Elaboration	Providing details about a previous statement	Student: I help patients. I give food, water. Is very importante
Elaboration request	Asking for more information or details	Student: Can you give more information?
Reply clarification	Student asks to explain a word or an idea based on a question made by AI.	AI: What was a challenge you faced during your previous job? Student: Challenge? No understand.

Reply confirmation	Student confirms their understanding by saying "yes".	AI: So you told me you have been working as a cashier for five years? Student: Yes
Reply elaboration	Providing the listener with more explanation after a question	AI: What would you do if a customer does not like your service? Student: I am patient, I talk with my boss and try to help.
Vocabulary check	Student checks if AI knows the correct word	Student: In my job, I use the machine for money. You know? AI: You mean "cash register"? Student: Ah yes.

Reflection Interviews

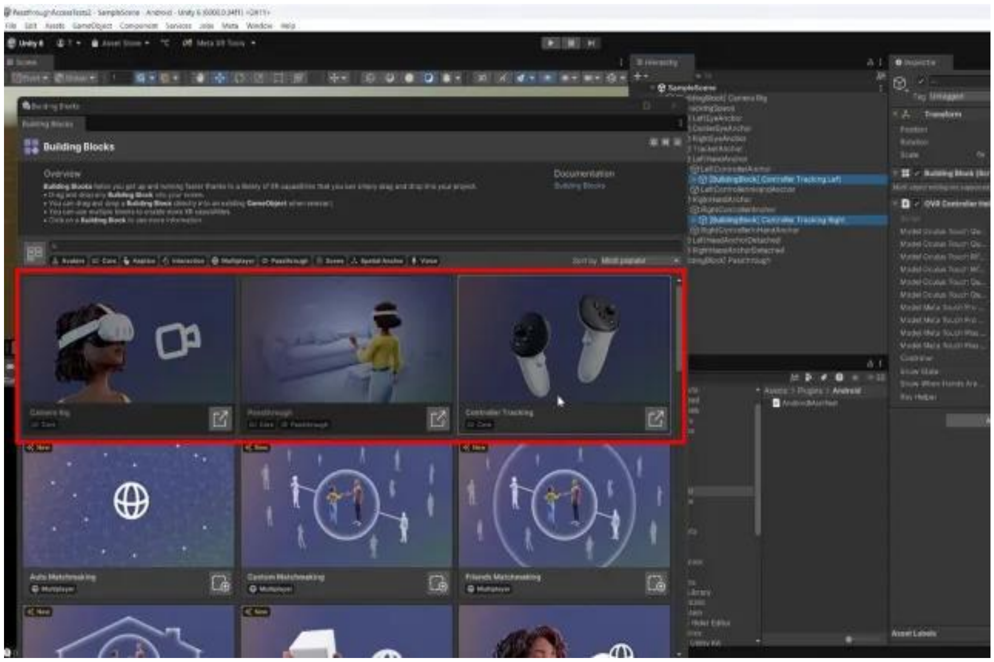
To integrate the findings for research question 4, this study employed thematic analysis following Braun and Clark's (2006) six-phase process, which provides a structured and rigorous method for identifying, analyzing, and presenting recurring themes in qualitative data. The thematic analysis involved three raters who coded 55 interviews with the three groups of participants. The researcher familiarized the raters with the study objectives, methodology, and reflection interview questions. A sample interview from all three groups was tested to determine whether the raters would reach an agreement. All the raters were experienced in language learning education. The study was approved by the Institutional Review Board (IRB) committee. Their teaching experience ranged from 10 to 15 years. Initial coding was conducted by systematically examining each student's responses and comparing the results to reach an agreement in identifying and labeling codes relevant to answering research question 4. The codes were then conducted using NVIVO 15 software and were organized into themes through grouping related to students' experiences with their assigned conditions, teachers' role, teaching method, and any emerging themes. Fourth, the themes were reviewed for internal coherence, and any disagreements between raters were resolved by discussing differences and similarities and reaching a consensus among raters. Fifth, each theme was labeled to reflect its central and distinctive ideas. Finally, the themes were organized into a logical and cohesive narrative to answer the research question. An interrater reliability of Fleiss k = 86 demonstrated consistent and reliable agreement between the three raters, in line with Landis and Koch (1977).

The Design of the Conversational AI Agent in MR Environment

The Mixed Reality environment was developed using the authoring tool Unity and the Meta interaction SDK. The SDK included the Meta Building Block pre-configured functionality that includes the passthrough feature that enables users to see the physical environment as well as the digital environment blended within it, as shown in Figure 3.

Meta's Building Block pre-configured functionality also includes 'scene understanding' that identifies physical objects such as walls, floors, and furniture to place virtual elements realistically. The scene included a virtual chair and a desk with 'ungrabbable'' or non-interactable objects placed on the desk, such as books and water bottles. The reason for not making these objects grabbable is to follow human-computer interaction best practices in terms of cognitive load and distraction (Dede, 2009; Radianti et al., 2020).

Figure 2 *Meta Building Block and the Interaction SDK Interface*



The MECAI used as the trainer in MR was built using Convai: https://convai.com/. Convai is a platform by NVIDIA that integrates ECAs and NVIDIA's LLMs for different purposes, including education. The MECAI was designed using the Ready Player Me platform, which is integrated through Convai. Convai allows for the design of the persona, tone, emotions, and knowledge bank of the MECAI. The Convai SDK was imported into Unity and set up using the Convai API key. The AI agent was first designed on Convai (Figure 4) using a Character Creator Tool. Then, using the Convai Character Importer, the character's prefab, or reusable 3D asset, was added to the Unity scene. The character was coded using C# for seating, gaze, and gestures. To follow best practices in conversational AI design, the conversation bubble that is usually on top of the characters heads where users can read the text of the conversation was removed to avoid distraction (Gonzales et al., 2025) and to align with the second experimental group where the learners communicated with other AI tools and did not see any text.

Figure 3

Convai Avatar Design Interface

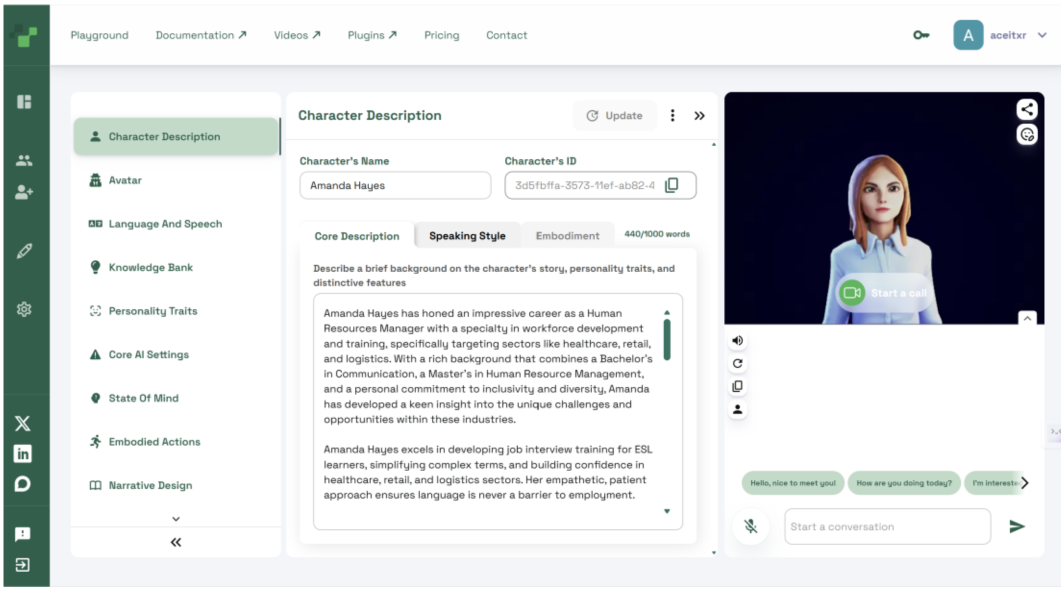
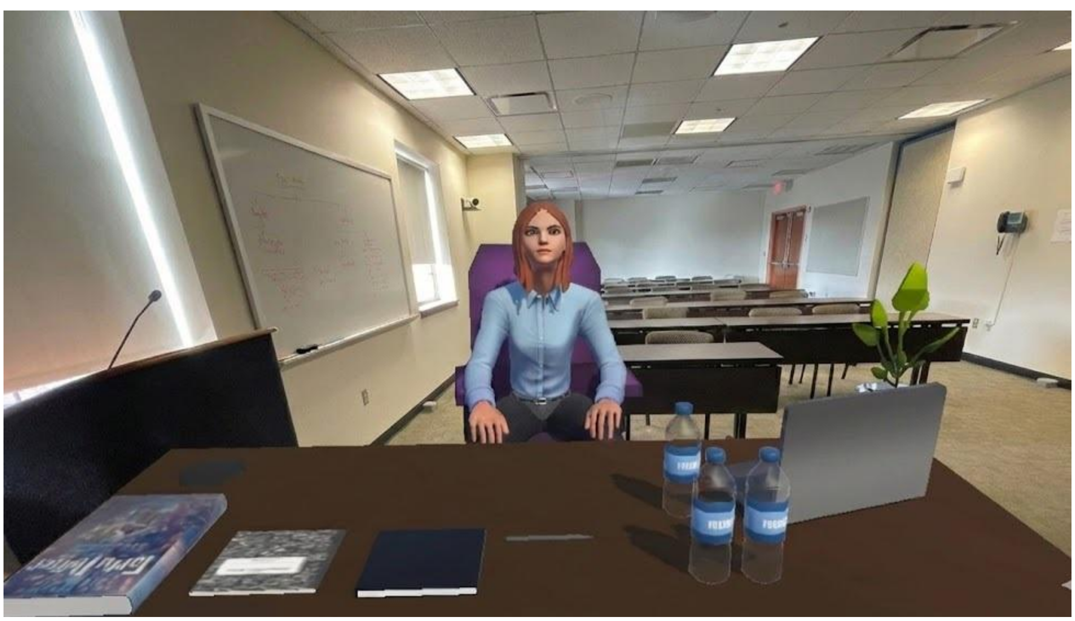


Figure 4

The Immersive Environment of the MR Simulation placed in a Physical classroom at HCCC



When students wear the HMD, they see MECAI in the classroom. They were trained to use the right controller to interact with the AI agent by pressing the A button to initiate speaking or interrupt the AI agent, as shown in Fig. 5.

Figure 5

Meta Quest Pro Controllers Mechanism for Initiating Student- MECAI Interaction in the MR Environment

ChatGPT Prompt Design

To enable ChatGPT to communicate with students as a job interview trainer, a prompt was created based on conversation design approaches in language learning contexts (Mabrito, 2024). The prompt provided ChatGPT with context, constraints, and guidelines to generate an output that aligned with the students' expectations and needs. Each student was oriented to its use and how they could upload the conversation logs into the classroom's Learning Management System (LMS) Canvas. The researcher also used Prompt Engineering GPT to modify the final prompt. The prompt was posted on the students' assignment page on Canvas. The prompt is shown in Table 6.

Table 6

ChatGPT prompt for the ChatGPT Voice group

You are my ESL teacher helping me prepare for job interview questions in English. I want you to: 1. Ask me one question at a time from this list:

- Tell me about yourself.
- Walk me through your resume.
- Why should we hire you?
- What can you bring to the company?
- What are your greatest strengths?
- What do you consider to be your weaknesses?
- What is your greatest professional achievement?
- Tell me about a challenge or conflict you've faced at work, and how you dealt with it.

2. Let me answer in my own words (sometimes with mistakes).
3. Give me constructive feedback: correct grammar, vocabulary, and fluency, but be brief and clear.
4. Suggest a better way to say my answer in professional English, while keeping it natural and simple.
5. Adjust your language to my English level when explaining

Ethical Considerations

In this study, all procedures adhered to the Teachers College Institutional Review Board (IRB) ethical guidelines. Letters of support were obtained from the HCCC. IRB approval was obtained before conducting the study. Students and external examiners signed an informed consent form that included detailed information about the study's aim, procedures, and participants' rights. The informed consent forms clarified that participation was voluntary and that participants could discontinue their involvement in the study at any time without any consequences.

To ensure equitable access to the conversational AI tools, following the study period, all participants were given the opportunity to experiment with a conversational AI tool that was not in their group. The study was conducted in a Skills for Success (Level 2) classroom that incorporated job interview skills as a component of the curriculum. The researcher also observed signs of discomfort from using the HMD. The students in the MR group did not experience dizziness or motion sickness. Students were advised to pause or discontinue the interaction process with the MECAI in MR if they felt discomfort or heaviness from the HMD. The primary concern in this study was the students' well-being, which was prioritized over data collection.

Issues of Validity and Reliability

Validity in mixed methods research refer to the "determination that an indicator actually measures what it is supposed to measure." (Abowitz & Toole, 2010, p. 8). Failure to do so is likely to result in bias in the resulting data. The main constructs of this study were speaking performance and speaking anxiety. This study consists of quantitative measures: pretests and posttests comprising eight mock job interview questions, FLCAS self-report that consists of 33 items on a 5- point Likert scale, and physiological measure using heart rate monitors to understand how different types of conversational AI tools can improve students' performance and reduce their anxiety (Følstad et al., 2018, October). Triangulating subjective (FLCAS) and objective measures (Heart Rate) is vital for accurate results (Ganster et al., 2018; Hardy, 1999; Schmidt et al., 2020; Teigen, 1994).

Speaking Performance

Speaking performance outcome measures were adapted from Ramanarayanan et al. 's(2017) analytic scoring rubric that measured dialogic interactions in language learning through three criteria: fluency, pronunciation, intonation, and stress. Although their rubric is not fully validated, their study established its practicality of their rubric in assessing dialogic interactions through interrater reliability. Accordingly, this study expanded the assessment by adding four more criteria aligned with job interview topics. This study also established interrater reliability for all seven criteria, which included the three criteria based on the previous study: fluency, pronunciation, and intonation and stress, in addition to the added criteria: grammar and sentence structure, vocabulary use, content and relevance of answers, and use of work-based examples. The interrater reliability of the seven criteria was employed by three experienced raters, including the researcher and two experienced raters. A substantial agreement among raters was established according to the guidelines of Landis and Koch (1977) with k = 75.

Speaking Anxiety

Speaking anxiety was assessed using two measures: a subjective measure, the Horwitz et al. (1986) FLCAS, and the HR monitor Polar OH1. In terms of reliability and validity, Horwitz's (1986) highly cited paper provided preliminary evidence for the FLCAS's validity and reliability. The results showed high internal consistency through Cronbach's alpha (α =.93). As for validity, the results show strong content and construct validity, linking item content to foreign language anxiety symptoms and justifying construct validity by indicating how the items relate to unique anxiety triggers. A considerable number of studies (Botes et al., 2020) observed high internal consistency of α > .90 (Aida, 1994; Elkhafaifi, 2005; Göcer, 2014). Moreover, Horwitz et al. (1986) noted the consistency of test-retest reliability of rtt = .83.

1The Polar OH1+ heart monitor showed high accuracy in different activity modes, ranging from sedentary to highly active tasks, including a correlation coefficient of 1.00 and a mean bias of 0.05 BPM (Schweizer & Gilgen-Ammann, 2025). This study aligns with established procedures in prior research in terms of measuring the participants' baseline followed by the task, the mock job interview, to separate normal arousal from task-induced anxiety so that a clear interpretation of anxiety can be obtained (Kao et al., 2014; Lehrer et al., 2007).

Triangulation and Qualitative Validity

To enhance the validity of the study findings, two qualitative data methods were employed: the analysis of the six-week intervention and the reflection interviews following the intervention. Thus, gaining rich insights into the students patterns of interactions with each of the conversational AI agents, and gaining rich, detailed account of their perception of the whole process including condition assignment, teaching method, teacher role, and their perceptions about how AI improved their speaking and lowered their anxiety.

To gain a closer look at how interaction with AI helped the students improve their speaking performance, examining the patterns of negotiation of meaning during the six-week interaction was also employed in this study. NoM is identified as the way a learner improves by moving from input – interaction – modified output. Long (1983) viewed NoM as helping learners improve their communicative skills by facilitating comprehensible output (i.e., language that the learner understands and aligns with their level of proficiency). Research has also established that learners who negotiate more learn more (Gass & Mackey, 2007; Pika, 1994). Researchers have also validated coding categories and taxonomies (Patterson and Trabaldo, 2006). As established by research, the interrater reliability of coding considers validated categories and taxonomies. In this study, two coders were involved in the percentage agreement, aligning with Samani et al. (2015). Raters agreed on 85& of their coding, which revealed substantial agreement between coders.

For the reflection interviews, students were asked 11 questions (Appendix A) that ranged from questions about their experience with their condition, how it impacted their learning, anxiety, and how the teacher's role impacted the whole process. The researcher was keen on asking open-ended questions that did not lead to any specific responses and that could capture the utmost experience that the students had. To ensure the reliability and consistency of coding, three independent raters, including the researcher, who explained the purpose of the study and the procedure, calculated the interrater reliability after coding the interviews on SPSS 15. Fleiss' kappa coefficient of 0.74 indicates substantial agreement among the raters.

Summary

This methodology chapter describes the research design of this study, the data collection employed to inform the results, and the established practices to validate these results. The study's sample population consisted of 55 students (N= 55). Their age range is between 18-64 years old. They were college-level ESL students at Hudson County Community College in New Jersey, enrolled in the Skills for Success 2 course. The students were randomly assigned to three conditions (control group, ChatGPT voice app group, and MECAI in MR group) following the pretest. The experimental group was involved in a six-week intervention, where they interacted with the AI tools for 20 minutes, whereas the control group received no treatment and was involved in classroom-based activities. Following the intervention and posttests, the students were interviewed to obtain their perceptions of the study. This study aims to fill a gap in language acquisition research and contribute to the growing body of research employing embodied conversational AI in immersive environments for language acquisition.

Chapter IV

Results

Introduction

Quantitative data were collected in the form of pre- and posttest measures that examined speaking performance and anxiety. For speaking performance, mock job interview questions were asked before and after the intervention. For speaking anxiety, FLCAS self-reports and heart rates were administered before and after the intervention. To examine how interaction with AI improves language acquisition, NoM patterns were analyzed. Following the intervention and posttest, reflection interviews were conducted with the participants.

Overview of the data analysis

Data were analyzed using SPSS version 23.0 for Windows (SPSS Inc., Chicago, USA). The normality of continuous data was tested using the Kolmogrov-Smirnov and Shapiro-Wilk tests. Continuous data are described as mean, standard deviation (SD), median, interquartile range (IQR), and minimum and maximum values. Categorical data were described by frequency and percentage.

Differences between the study groups (Traditional, ChatGPT voice app, and MECAI in MR) in the pre- and post-intervention scores of the seven language performance domains (fluency, pronunciation, intonation and stress, grammar and sentence structure, vocabulary use, content and relevance of answers, and use of work-based examples) were assessed using the chi-squared test with Monte Carlo correction. Differences in the pre- and post-intervention scores of the language performance domains for each study group were assessed using the Wilcoxon signed-rank test. Seven ordinal logistic regression models were created to assess the post-intervention scores of the language performance domains through the study groups while controlling for the pre-intervention scores.

The difference between the study groups in the pre- and post-intervention total language performance scores was tested using One-way ANOVA with post hoc Bonferroni adjustment. The difference in each study group between the pre- and post-intervention total language performance scores was tested using a paired t-test. Repeated measures ANOVA with post hoc Bonferroni adjustment and a linear regression model were created to assess the post-intervention total language performance score through the study groups, controlling for the pre-intervention score and the interaction between study groups and pre-intervention scores.

Differences between the study groups in the pre-interview, post-interview, and during-interview heart rates were tested using One-way ANOVA. Differences in each study group between the pre-interview and post-interview baseline and during-interview heart rates were tested using a paired t-test. Repeated measures of ANOVA tests with two linear regression models were created to assess the post-interview baseline and during-interview heart rates through the study groups, controlling for the pre-interview heart rate and interaction between study groups and pre-interview heart rate.

The difference between the study groups in the FLCAS was tested using a One-way ANOVA. The difference in each study group between the pre- and post-intervention FLCAS was tested using a paired t-test. Repeated measures of ANOVA with a linear regression model were created to assess the post-intervention FLCAS through the study groups, controlling for the pre-intervention FLCAS and the interaction between study groups and pre-intervention FLCAS.

X, Z, F, and t values, partial Eta squared ($\eta 2$), Nagelkerke R^2, adjusted R^2, adjusted odds ratios, B regression coefficients, and 95% confidence intervals were calculated. p value <0.05 was considered statistically significant.

Description of the study sample

Fifty-five students participated in this study. The majority of the students were females (76.36%). The mean (SD) of age = 29.20 (13.65). The mean (SD) of their stay duration in the USA= 37.13 (46.23). The majority (87.27%) did not speak English at home and had previously applied for jobs in the USA (74.55%). Students from the Dominican Republic represented approximately one-third of the study sample (27.28%), followed by Ecuador (21.82%) and Egypt (12.73%). Approximately 40% were neutrally confident when speaking English. Over half the sample (60%) had never received formal training in English. The mean (SD) number of months of learning English was 8.95 (13.87) (Table 7).

The mean (SD) of the pre-intervention total language performance score, post-intervention total language performance score, pre-intervention total FLCAS, post-intervention total FLCAS, Pre-interview baseline heart rate, post-interview baseline heart rate, pre-interview during-interview heart rate, and post-interview during-interview heart rate were 8.58 (2.56), 15.95 (3.15), 89.24 (16.84), 89.07 (19.27), 89.65 (15.48), 83.49 (15.39), 86.41(12.16), and 83.96 (16.59), respectively (Table 7).

Table 7

Description of the study sample (N=55)

Age (years)	**Mean (SD)**	29.20 (13.65)
	Median (IQR)	23.00 (19.00, 40.00)
	Min-Max	17.00 – 63.00
Gender	**Females n (%)**	42 (76.36)
	Males n (%)	13 (23.64)
Duration of residence in the USA (months)	**Mean (SD)**	37.13 (46.23)
	Median (IQR)	24.00 (8.00, 42.00)
	Min-Max	2.00 – 240.00
Speaking English at home	**Yes n (%)**	7 (12.73)
	No n (%)	48 (87.27)
Country of origin	**Afghanistan n (%)**	1 (1.82)
	Algeria n (%)	2 (3.64)

	Argentina n (%)	1 (1.82)
	Colombia n (%)	3 (5.45)
	Dominican Republic n (%)	15 (27.28)
	Ecuador n (%)	12 (21.82)
	Egypt n (%)	7 (12.73)
	India n (%)	1 (1.82)
	Mexico n (%)	1 (1.82)
	Morocco n (%)	2 (3.64)
	Peru n (%)	6 (10.91)
	Senegal n (%)	2 (3.64)
	Sudan n (%)	1 (1.82)
	United States n (%)	1 (1.82)
Previously applied for job in the US	**Yes n (%)**	41 (74.55)
	No n (%)	14 (25.45)
Having confidence speaking English	**Very confident n (%)**	8 (10.91)
	Somewhat confident n (%)	17 (30.91)
	Neutral n (%)	21 (38.18)
	Somewhat unconfident n (%)	6 (10.91)
	Very unconfident n (%)	3 (5.45)
Previous formal training on English	**Yes n (%)**	22 (40.00)
	No n (%)	33 (60.00)
Number of months of learning English	**Mean (SD)**	8.95 (13.87)
	Median (IQR)	6.00 (3.00, 8.00)
	Min-Max	1.00 – 96.00
Pre-intervention total language performance score[a]	**Mean (SD)**	8.58 (2.56)
	Median (IQR)	7.00 (7.00, 9.00)
	Min-Max	7.00 – 18.00
Post-intervention total language performance score[a]	**Mean (SD)**	15.95 (3.15)
	Median (IQR)	17.00 (14.00, 18.00)
	Min-Max	7.00 – 21.00
Pre-intervention total FLCAS[b]	**Mean (SD)**	89.24 (16.84)
	Median (IQR)	90.00 (78.00, 99.00)
	Min-Max	45.00 – 145.00
Post-intervention total FLCAS[b]	**Mean (SD)**	89.07 (19.27)
	Median (IQR)	90.00 (80.00, 103.00)
	Min-Max	46.00 – 147.00
Pre-interview baseline heart rate (beats per minute)	**Mean (SD)**	89.65 (15.48)
	Median (IQR)	90.46 (76.32, 104.80)
	Min-Max	60.56 – 116.33
Post-interview baseline heart rate (beats per minute)	**Mean (SD)**	83.49 (15.39)
	Median (IQR)	82.97 (71.23, 96.25)

	Min-Max	51.95 – 117.40
Pre-interview during-interview heart rate (beats per minute)	**Mean (SD)**	86.41 (12.16)
	Median (IQR)	86.39 (78.14, 94.52)
	Min-Max	60.56 – 117.58
Post-interview during-interview heart rate (beats per minute)	**Mean (SD)**	83.96 (16.59)
	Median (IQR)	87.23 (71.17, 96.56)
	Min-Max	43.71 – 116.96

SD Standard deviation, IQR Interquartile range, Min Minimum, Max Maximum, FLCAS Foreign Language Classroom Anxiety Scale

[a] Language performance scale ranges from 7 (lowest) to 21 (highest)

[b] FLCAS ranges from 33 (lowest) to 165 (highest)

Analysis for Research Question 1 (Speaking Performance)

There was no statistically significant difference between the study groups in the pre-intervention fluency score (X=5.28, p=0.20). There was a statistically significant difference between the study groups in the post-intervention fluency score, where the percentage of students who scored excellent in the mixed reality AI group (80.00%) was higher than that in the traditional (13.3%) and ChatGPT voice app (35.0%) groups (X=21.41, p=<0.001). There was a statistically significant difference in pre- and post-intervention fluency scores in all groups (p<0.05) (Table 8, Figure 6).

Table 8

Difference between study groups in fluency score pre-and post-intervention (N=55)

Fluency score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	p value [a]
Pre-intervention	**Needs improvement**	14 (93.33)	17 (85.00)	14 (70.00)	5.28	0.20
	Satisfactory	1 (6.67)	2 (10.00)	6 (30.00)		
	Excellent	0 (0.00)	1 (5.00)	0 (0.00)		
Post-intervention	**Needs improvement**	4 (26.67)	0 (0.00)	0 (0.00)	21.41	<0.001*
	Satisfactory	9 (60.00)	13 (65.00)	4 (20.00)		
	Excellent	2 (13.33)	7 (35.00)	16 (80.00)		
***p* value [b]**		0.001*	<0.001*	<0.001*	-	-
Z		3.21	4.06	3.87		

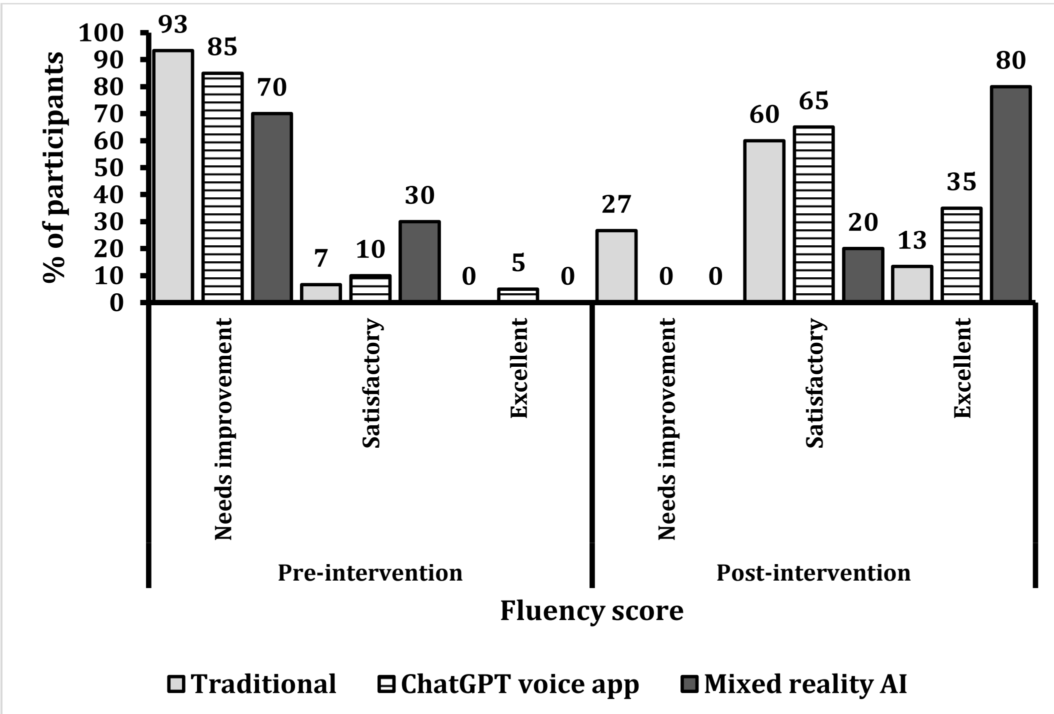
CI Confidence interval

[a]Chi-squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 6

Difference between study groups in fluency score pre-and post-intervention



In ordinal logistic regression, the ChatGPT voice app group had significantly higher odds of being in a higher category of post-intervention fluency score than the traditional learning group (AOR = 6.32, 95% CI: 1.11, 35.88, *p*= 0.04). Likewise, the mixed reality AI group had significantly higher odds of being in a higher category of the post-intervention fluency score compared with the traditional learning group (AOR = 41.93, 95% CI: 6.00, 293.14, *p*=<0.001). The regression model predicted 19% of the variance in post-intervention fluency scores. The model explained 19% of the variance in the post-intervention fluency score (Table 9).

Table 9

Ordinal logistic regression of the difference between the study groups in post-intervention fluency scores when controlling pre- intervention scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	6.32 (1.11, 35.88)	0.89	0.04*
Mixed reality AI	41.93 (6.00, 293.14)	0.99	<0.001*

Nagelkerke $R^2 = 0.19$
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant $p<0.05$

There was no statistically significant difference between the study groups in the pre-intervention pronunciation scores (X=0.95, *p*=0.60). There was a statistically significant difference between the study groups in the post-intervention fluency score, where the percentage of students who scored excellent in the mixed reality AI group (40.00%) was higher than the traditional (0.00%), and ChatGPT voice app (20.00%), (X=11.76, *p*=0.01). There was a statistically significant difference in the pre- and post-intervention pronunciation scores in all groups (*p*<0.05) (Table 10, Figure 7).

Table 10

Difference between study groups in pronunciation scores pre-and post-intervention (N=55)

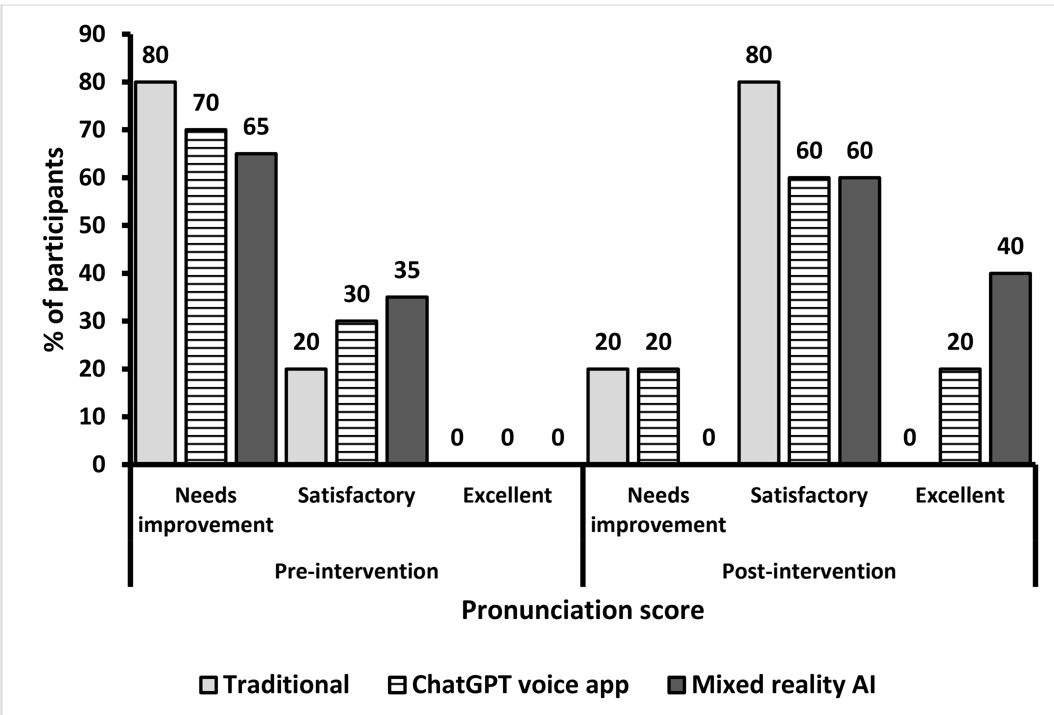
Pronunciation Score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value [a]
Pre-intervention	**Needs improvement**	12 (80.00)	14 (70.00)	13 (65.00)	0.95	0.60
	Satisfactory	3 (20.00)	6 (30.00)	7 (35.00)		
	Excellent	0 (0.00)	0 (0.00)	0 (0.00)		
Post-intervention	**Needs improvement**	3 (20.00)	4 (20.00)	0 (0.00)	11.76	0.01*
	Satisfactory	12 (80.00)	12 (60.00)	12 (60.00)		

	Excellent	0 (0.00)	4 (20.00)	8 (40.00)		
***p* value** [b]		0.003*	0.005*	<0.001*	-	-
Z		3.00	2.84	3.83		

CI Confidence interval
[a]Chi- squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test
*Statistically significant at p <0.05

Figure 7

Difference Between Study Groups In Stress And Annotation Score Pre-and Post-Intervention



In ordinal logistic regression, the ChatGPT voice app group was not significantly associated with the post-intervention pronunciation score (AOR = 2.19, 95% CI: 0.48, 9. The MECAI group had significantly higher odds of being in a higher category of the post-intervention pronunciation score compared with the traditional learning group (AOR = 10.36, 95% CI: 2.05, 52.33, *p*= 0.005).The model explained 3% of the variance in the post-intervention pronunciation score (Table 11).

Table 11

Ordinal logistic regression of the association between the study groups in post-intervention pronunciation scores when controlling pre- intervention scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	2.19 (0.48, 9.96)	0.77	0.31
Mixed reality AI	10.36 (2.05, 52.33)	0.83	0.005*

Nagelkerke $R^2 = 0.03$
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant $p<0.05$

There was no statistically significant difference between the study groups in pre-intervention stress and annotation scores (X=1.86, p=0.88). There was a statistically significant difference between the study groups in the post-intervention stress and annotation score, where the percentage of students who scored excellent in the mixed reality AI group (40.00%) was higher than the traditional (6.67%), and ChatGPT voice app (25.00%), (X=8.89, p=0.049). There was a statistically significant difference in all groups between the pre- and post-intervention stress and annotation scores (p<0.05) (Table 12, Figure 8).

Table 12

Difference between study groups in stress and annotation score pre-and post-intervention (N=55)

Stress and annotation score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value [a]
Pre-intervention	**Needs improvement**	13 (86.67)	17 (85.00)	15 (75.00)	1.86	0.88
	Satisfactory	2 (13.33)	2 (10.00)	4 (20.00)		
	Excellent	0 (0.00)	1 (5.00)	1 (5.00)		
Post-intervention	**Needs improvement**	4 (26.67)	3 (15.00)	0 (0.00)	8.89	0.049*
	Satisfactory	10 (66.67)	12 (60.00)	12 (60.00)		

	Excellent	1 (6.67)	5 (25.00)	8 (40.00)		
***p* value** [b]		0.002*	<0.001*	<0.001*	-	-
X		3.16	3.63	3.64		

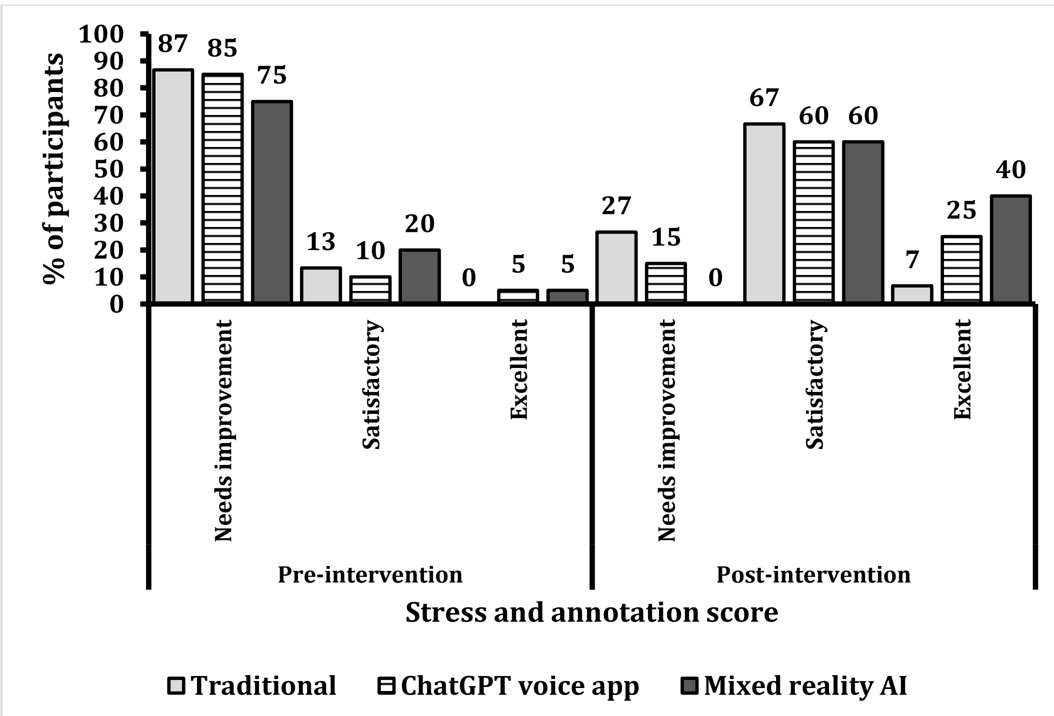
CI Confidence interval

[a]Chi-squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 8

Difference between study groups in stress and annotation score pre-and post-intervention



In ordinal logistic regression, the ChatGPT voice app group was not significantly associated with post-intervention stress and annotation score (AOR = 2.81, 95% CI: 0.62, 12.71, p= 0.18). The mixed reality AI group had significantly higher odds of being in a higher category of post-intervention stress and annotation score compared with the traditional learning group (AOR = 8.12, 95% CI: 1.66, 39.85, p= 0.01). The model explained 16% of the variance in post-intervention pronunciation scores (Table 13).

Table 13

Ordinal logistic regression of the association between the study groups and post-intervention stress and annotation scores when controlling pre-interview scores (N=55)

Explanatory variable	AOR (95% CI)	Standard Error	*p* value
Traditional	Reference category		
ChatGPT voice app	2.81 (0.62, 12.71)	0.77	0.18
Mixed reality AI	8.12 (1.66, 39.85)	0.81	0.01*

Nagelkerke $R^2 = 0.16$
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant $p < 0.05$

There was no statistically significant difference between the study groups in the pre-intervention grammar and sentence structure scores (X=2.42, p=0.88). There was a statistically significant difference between the study groups in the post-intervention grammar and sentence structure scores, where the percentage of students who scored excellent in the mixed reality AI group (70.00%) was higher than in the traditional (13.33%) and ChatGPT voice app (35.00%) groups (X=12.65, p= 0.004).There was a statistically significant difference in all groups between the pre- and post-intervention grammar and sentence structure scores (p <0.05) (Table 14, Figure 9).

Table 14

Difference between study groups in grammar and sentence structure score pre-and post-intervention (N=55)

Grammar and sentence structure score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value [a]
Pre-intervention	**Needs improvement**	11 (73.33)	16 (80.00)	16 (80.00)	2.42	0.88
	Satisfactory	4 (26.67)	3 (15.00)	4 (20.00)		
	Excellent	0 (0.00)	1 (5.00)	0 (0.00)		
Post-intervention	**Needs improvement**	2 (13.33)	1 (5.00)	0 (0.00)	12.65	0.004*

	Satisfactory	11 (73.33)	12 (60.00)	6 (30.00)		
	Excellent	2 (13.33)	7 (35.00)	14 (70.00)		
***p* value** [b]		0.005*	<0.001*	<0.001*		-
Z		2.84	3.83	3.95		

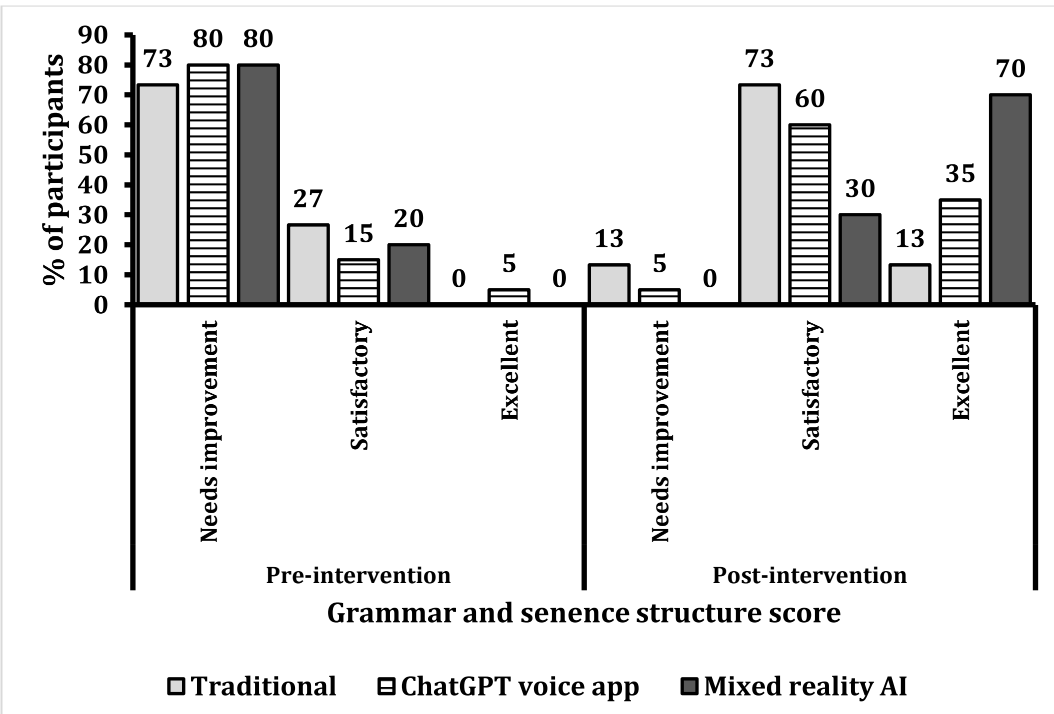
CI Confidence interval

[a]Chi-squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 9

Difference between study groups in grammar and sentence structure score pre-and post-intervention



In ordinal logistic regression, the ChatGPT voice app group was not significantly associated with post-intervention grammar and sentence structure scores (AOR = 3.27, 95% CI: 0.68, 15.79, p= 0.14). Likewise, the mixed reality AI group had significantly higher odds of being in a higher category of the post-intervention fluency score compared with the traditional learning group (AOR = 41.93, 95% CI: 6.00, 293.14, p=<0.001). The regression model predicted 19% of the variance in post-intervention fluency scores. The model explained 19% of the variance in the post-intervention fluency score (Table 9).

Table 15

Ordinal logistic regression of the association between the study groups and post-intervention grammar and sentence structure scores when controlling pre-intervention scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	3.27 (0.68, 15.79)	0.80	0.14
Mixed reality AI	17.10 (3.24, 90.19)	0.85	<0.001*

Nagelkerke R squared = 0.06
AOR Adjusted odds ratio, CI confidence interval, SE Standard error
*Statistically significant p <0.05

There was a statistically significant difference between the study groups in the pre-intervention vocabulary use score, (X=8.40, p=0.03). There was a statistically significant difference between the study groups in the post-intervention grammar and sentence structure scores, where the percentage of students who scored excellent in the mixed reality AI group (70.00%) was higher than in the traditional (13.33%) and ChatGPT voice app (35.00%) groups (X=12.65, p= 0.004). There was a statistically significant difference in all groups between the pre- and post-intervention grammar and sentence structure scores (p <0.05) (Table 14, Figure 9).

Table 16

Difference between study groups in vocabulary use scores pre-and post-intervention (N=55)

Vocabulary use score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value [a]
Pre-intervention	**Needs improvement**	14 (99.33)	17 (85.00)	12 (60.00)	8.40	0.03*
	Satisfactory	1 (6.67)	2 (10.00)	8 (40.00)		

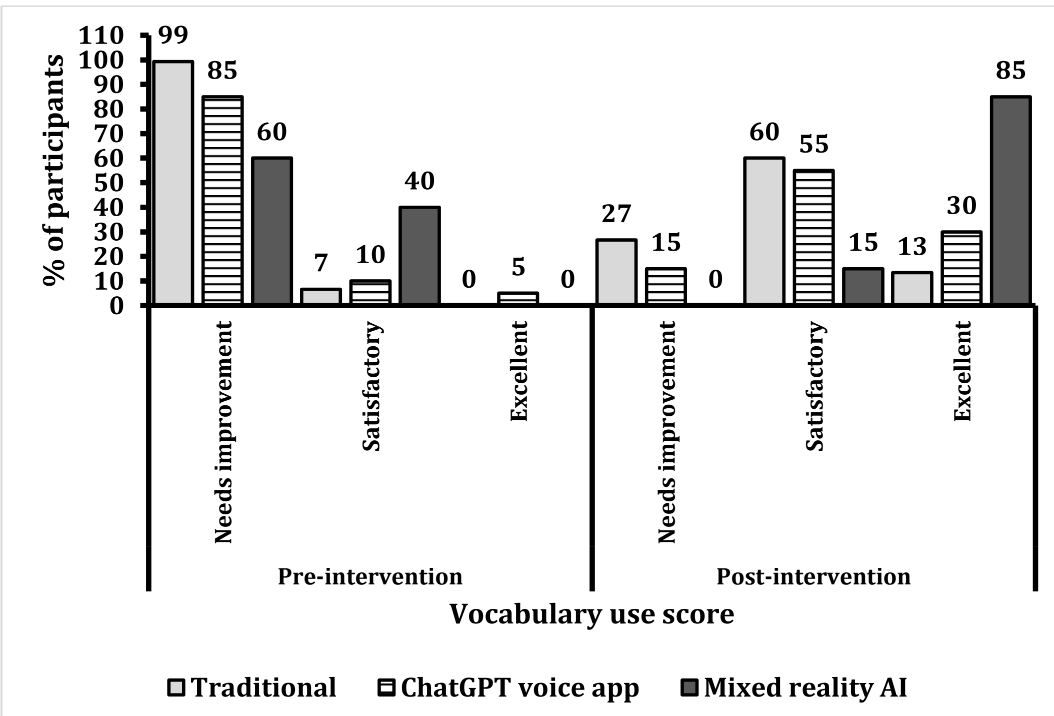
	Excellent	0 (0.00)	1 (5.00)	0 (0.00)		
Post-intervention	**Needs improvement**	4 (26.67)	3 (15.00)	0 (0.00)	21.64	<0.001*
	Satisfactory	9 (60.00)	11 (55.00)	3 (15.00)		
	Excellent	2 (13.33)	6 (30.00)	17 (85.00)		
***p* value [b]**		0.001*	<0.001*	<0.001*	-	-
X 2		3.21	3.58	4.04		

CI Confidence interval

[a]Chi-squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 10 *Difference between study groups in vocabulary use scores pre-and post-intervention*



In ordinal logistic regression, the ChatGPT voice app group was not significantly associated with the post-intervention vocabulary use score compared with the traditional group (AOR = 1.98, 95% CI: 0.48, 8.13, *p*= 0.34). The mixed reality AI group had significantly higher odds of being in a higher category of the post-intervention vocabulary use score than the traditional learning group (AOR = 28.20, 95% CI: 4.66, 170.66, *p*= <0.001). The model explained 28% of the variance in the post-intervention vocabulary-use score (Table 17).

Table 17

Ordinal logistic regression of the association between the study groups and post-intervention vocabulary use scores when controlling pre-interview scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	1.98 (0.48, 8.13)	0.72	0.34
Mixed reality AI	28.20 (4.66, 170.66)	0.92	<0.001*

Nagelkerke $R^2 = 0.28$
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant $p < 0.05$

There was no statistically significant difference between the study groups in the pre-intervention content and relevance of answers score, (X=5.21, *p*=0.13). There was a statistically significant difference between the study groups in the post-intervention content and relevance of answers score, where the percentage of students who scored excellent in the mixed reality AI group (75.00%) was higher than the traditional (13.33%) and ChatGPT voice app (50.00%) (X=20.63, *p*=<0.001). There was a statistically significant difference in all groups between the pre-intervention and post-intervention content and relevance of answers scores, (*p*=<0.05) (Table 18, Figure 11).

Table 18

Difference between study groups in content and relevance of answers score pre-and post-intervention (N=55)

Content and relevance of answers score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value [a]
Pre-intervention	**Needs improvement**	14 (93.33)	18 (90.00)	13 (65.00)	5.21	0.13
	Satisfactory	1 (6.67)	2 (10.00)	7 (35.00)		
	Excellent	0 (0.00)	0 (0.00)	0 (0.00)		
Post-intervention	**Needs improvement**	6 (40.00)	0 (0.00)	0 (0.00)	20.63	<0.001*

	Satisfactory	7 (46.67)	10 (50.00)	5 (25.00)		
	Excellent	2 (13.33)	10 (50.00)	15 (75.00)		
***p* value [b]**		0.004*	<0.001*	<0.001*	-	-
Z		2.89	3.84	3.94		

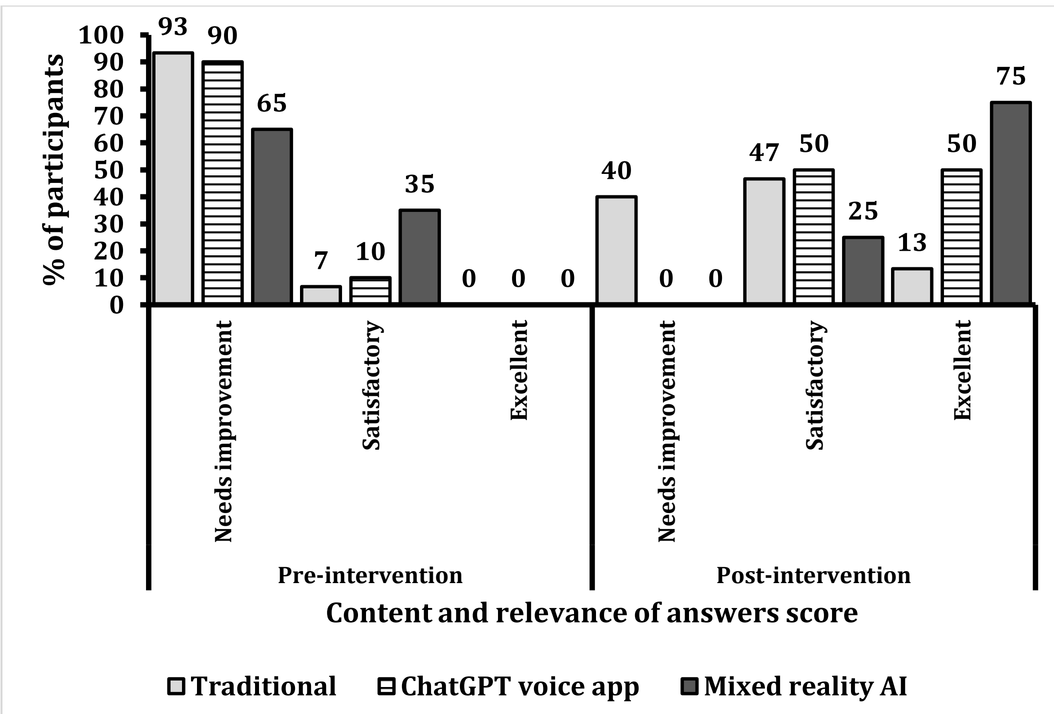
CI Confidence interval

[a]Chi- squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 11

Difference between study groups in content and relevance of answers score pre-and post-intervention



In ordinal logistic regression, the ChatGPT voice app group was non-significantly with the post-intervention content and relevance of answers score (AOR = 14.04, 95% CI: 2.47, 79.91, p= 0.07). The mixed reality AI group had significantly higher odds of being in a higher category of the post-intervention content and relevance of answers score compared with the traditional learning group (AOR = 35.71, 95% CI 5.33, 239.19, p=0.03). The model explained 21% of the variance in the post-intervention content and relevance of the answers score (Table 19).

Table 19

Ordinal logistic regression of the association between the study groups and the post-intervention content and relevance of answers scores when controlling pre-interview scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	14.04 (2.47, 79.91)	0.89	0.07
Mixed reality AI	35.71 (5.33, 239.19)	0.97	0.03*

Nagelkerke R^2 = 0.21
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant p <0.05

There was no statistically significant difference between the study groups in the pre-intervention use of work-based example score (X=0.88, p=0.78). There was a statistically significant difference between the study groups in the post-intervention use of work-based examples score, where the percentage of students who scored excellent in the mixed reality AI group (80.00%) was higher than that in the traditional (13.33%) and ChatGPT voice app (45.00%) groups (X=20.43, p=<0.001). There was a statistically significant difference in all groups between the pre- and post-intervention work-based example scores (p=<0.05) (Table 20, Figure 12).

Table 20

Difference between study groups in use of work-based examples of answers score pre-and post-intervention (N=55)

Use of work-based examples score		Traditional (n=15) n (%)	ChatGPT voice app (n=20) n (%)	Mixed reality AI (n=20) n (%)	X	*p* value[a]
Pre-intervention	**Needs improvement**	13 (86.67)	15 (75.00)	15 (75.00)	0.88	0.78
	Satisfactory	2 (13.33)	5 (25.00)	5 (25.00)		
	Excellent	0 (0.00)	0 (0.00)	0 (0.00)		
Post-intervention	**Needs improvement**	6 (40.00)	1 (5.00)	0 (0.00)	20.43	<0.001*

	Satisfactory	7 (46.67)	10 (50.00)	4 (20.00)		
	Excellent	2 (13.33)	9 (45.00)	16 (80.00)		
***p* value** [b]		0.01*	<0.001*	<0.001*	-	-
Z		2.50	3.76	4.04		

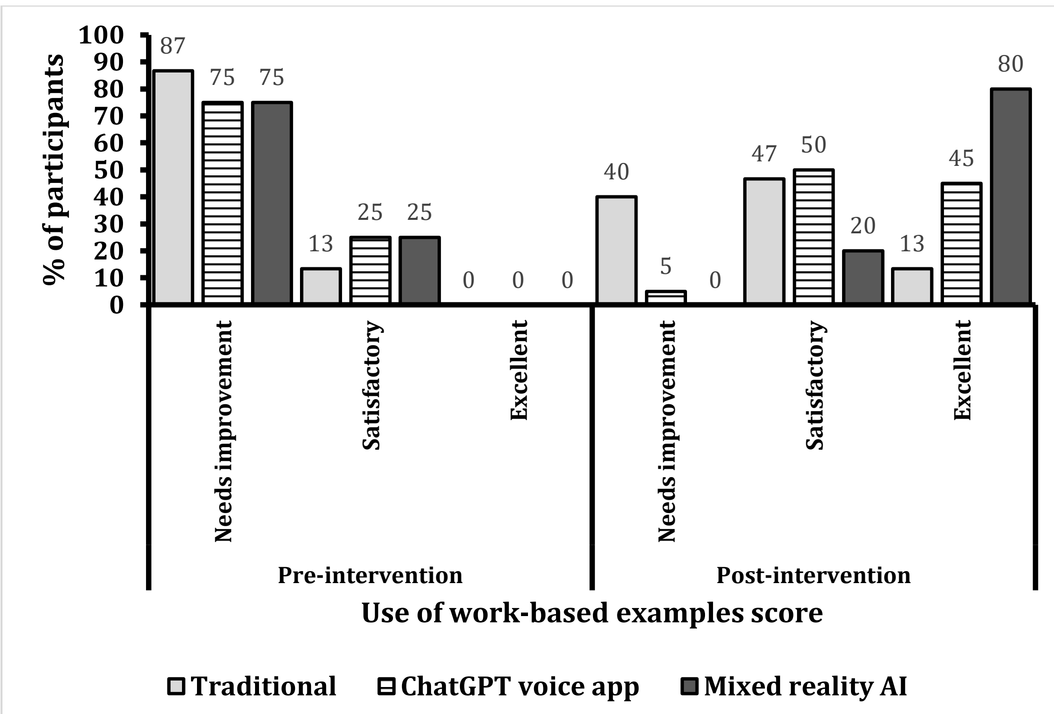
CI Confidence interval

[a]Chi- squared with Monte-Carlo correction, [b]Wilcoxon-signed rank test

*Statistically significant at p <0.05

Figure 12

Difference between study groups in use of work-based examples of answers score pre-and post-intervention



In ordinal logistic regression, the ChatGPT voice app group had significantly higher odds of being in a higher category of the post-intervention use of work-based examples score compared with the traditional learning group (AOR = 7.16, 95% CI: 1.52, 33.76, p=0.01). The mixed reality AI group had significantly higher odds of being in a higher category of the post-intervention use of work-based examples score compared with the traditional learning group (AOR = 37.27, 95% CI 6.38, 217.83, *p*=<0.001). The model explained 16% of the variance in the post-intervention work-based example score (Table 21).

Table 21

Ordinal logistic regression of the association between the study groups and post-intervention use of work-based examples scores when controlling pre-interview scores (N=55)

Explanatory variable	AOR (95% CI)	SE	*p* value
Traditional	Reference category		
ChatGPT voice app	7.16 (1.52, 33.76)	0.79	0.01*
MECAI in MR	37.27 (6.38, 217.83)	0.90	<0.001*

Nagelkerke R^2 = 0.16
AOR Adjusted odds ratio, CI confidence interval, SE standard error
*Statistically significant *p* <0.05

There was no statistically significant difference between the study groups in the pre-intervention total language performance score (partial η^2=0.04, F=0.98, *p*=0.38). There was a statistically significant difference between the study groups in the post-intervention total language performance score, where the mean (SD) score in the mixed reality AI group 18.40 (1.46) was higher than the traditional 12.87 (3.46), and ChatGPT voice app 15.80 (1.82) groups (Partial η^2=0.49, F=25.17, *p*=<0.001) (Table 22). The post hoc test showed a significant difference between all pairs of the study groups (*p*=<0.05) (Table 23). There was a statistically significant difference in all groups between the pre- and post-intervention total language performance scores (*p*=<0.001) (Table 22, Figure 13).

Table 22

Difference between study groups in the mean total language performance scores pre-and post-intervention (N=55)

Total language performance score	Traditional (n=15)	ChatGPT voice app (n=20)	Mixed reality AI (n=20)	Partial η^2	F	*p* value [a]

Pre-intervention	**Mean (SD)**	7.93 (1.91)	8.50 (2.78)	9.15 (2.74)	0.04	0.98	0.38
	Median (IQR)	7.00 (7.00, 8.00)	7.50 (7.00, 8.75)	8.00 (7.00, 10.75)			
	Min-Max	7.00 – 14.00	7.00 – 18.00	7.00 – 15.00			
Post-intervention	**Mean (SD)**	12.87 (3.46)	15.80 (1.82)	18.40 (1.46)	0.49	25.17	<0.001*
	Median (IQR)	14.00 (11.00, 15.00)	16.50 (14.00, 17.00)	18.00 (18.00, 19.00)			
	Min-Max	7.00 – 20.00	12.00 – 18.00	14.00 – 21.00			
Mean difference (SD)		-4.93 (3.33)	-7.30 (2.49)	-9.25 (2.55)	-		
t		-5.74	-13.09	-16.21			
95% CI		(-6.78, -3.09)	(-8.47, -6.13)	(-10.44, -8.06)			
***p* value [b]**		<0.001*	<0.001*	<0.001*			

SD Standard deviation, IQR Interquartile range, Min Minimum, Max Maximum, CI Confidence interval, η^2 Eta squared

[a]One-way ANOVA, [b]Paired t-test

*Statistically significant at $p<0.05$

Figure 13 *Difference between study groups in the mean total language performance scores pre- and post-intervention*

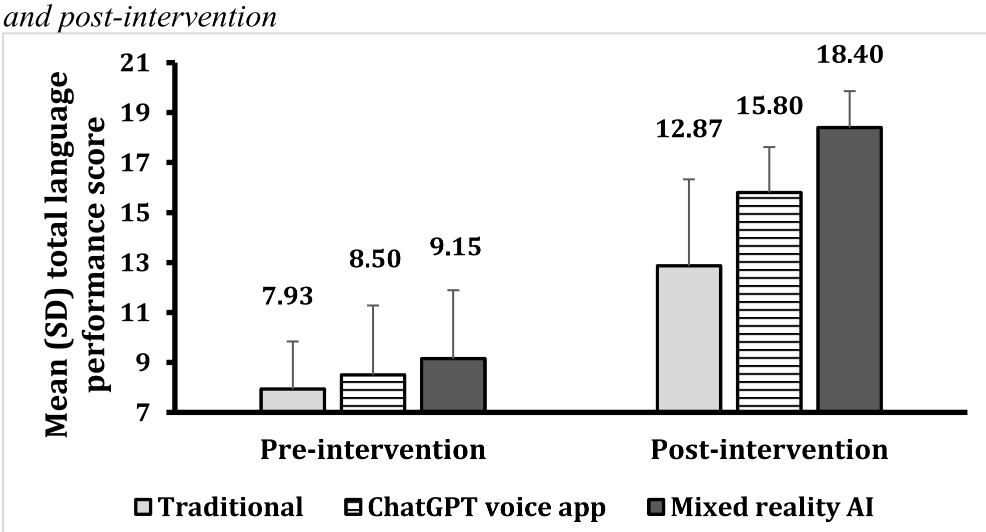


Table 23

Pairwise comparisons of the study groups in the post-intervention total score of language performance (N=55)

Comparison groups		Mean Difference	SE	Cohen's d	95% CI	*p* value
Traditional	**ChatGPT voice app**	-2.93	0.78	1.06	(-4.68, -1.00)	0.001*
	Mixed reality AI	-5.53	0.78	2.08	(-7.46, -3.60)	<0.001*
ChatGPT voice app	**Mixed reality AI**	-2.60	0.72	1.58	(-4.39, -0.81)	0.002*

SE Standard error, CI Confidence interval

Bonferroni adjustment

*Statistically significant at $p<0.05$

In linear regression, the ChatGPT voice app group was not significantly associated with the post-intervention total language performance score compared to the traditional group (B= 5.24, 95%CI: -0.22, 10.59, p=0.06). The mixed reality group had a significantly higher post-intervention total language performance score than the traditional group (B= 8.58, 95%CI: 2.99, 14.18, p=0.002). The pre-intervention total language performance score was significantly associated with a higher post-intervention total language performance score (B= 0.63, 95%CI: 0.06, 1.19, p=0.03). The interactions between the ChatGPT voice app and pre-intervention score, as well as between mixed reality AI and pre-intervention score, were not significantly associated with post-intervention total language performance score (B= -0.31, 95%CI: -0.97, 0.34, p=0.35) and (B= -0.42, 95%CI: -1.07, 0.24, p=0.21) (Table 24).

Table 24

Linear regression of the association between post-intervention total language performance score, study groups, and pre-intervention scores (N=55)

Explanatory variables		B (95% CI)	SE	*p* value
Group	**Traditional**	Reference category		
	ChatGPT voice app	5.24 (-0.22, 10.59)	2.78	0.06

	Mixed reality AI	8.58 (2.99, 14.18)	2.86	0.002*
Pre-intervention total language performance score		0.63 (0.06, 1.19)	0.29	0.03*
Traditional * pre-intervention score		Reference category		
ChatGPT voice app * pre-intervention score		-0.31 (-0.97, 0.34)	0.33	0.35
Mixed reality AI * pre-intervention score		-0.42 (-1.07, 0.24)	0.33	0.21

Adjusted R^2 = 0.524
B Regression coefficient, SE Standard error
*Statistically significant at p<0.05

In repeated measures of ANOVA, there was a significant association between post-intervention total language performance score and study groups, (Partial η^2 = 0.98, p=<0.001) (Table 25). Pairwise comparison showed that there were significant associations between each pair of groups (p=<0.05) (Table 26). The post-intervention total language performance score was significantly associated with the interaction between the study groups and pre-intervention total language performance score and (Partial η^2 = 0.29, p=<0.001) (Table 25).

Table 25

Association between post-intervention total language performance scores, study groups, and the interaction between the study groups and the pre-intervention scores (N=55)

Variables	**F**	**Partial η^2**	***p* value**
Groups	12.32	0.98	<0.001*
Groups * pre-intervention score	10.47	0.29	<0.001*

η^2 Eta squared
Repeated measures of ANOVA
*Statistically significant at p<0.05

Table 26

Pairwise comparisons of the study groups in the post-intervention total language performance scores (N=55)

Comparison groups		**Mean Difference**	**SE**	**95% CI**	***p* value**
Traditional	**ChatGPT voice app**	-1.75	0.68	(-3.44, -0.06)	0.04*
	Mixed reality AI	-3.38	0.68	(-5.06, -1.69)	<0.001*

ChatGPT voice app	**Mixed reality AI**	-1.63	0.63	(-3.19, -0.06)	0.04*

SE Standard error, CI Confidence interval
Bonferroni adjustment
*Statistically significant at *p* <0.05

Analysis for Research Question 2 (Anxiety and Heart Rate)

There was no statistically significant difference between the study groups in the pre-interview baseline heart rate, where the mean (SD) pre-interview baseline heart rates in the traditional, ChatGPT voice app, and mixed reality AI groups were 93.89 (16.25), 87.26 (16.09), and 88.86 (14.39), respectively (Partial η^2=0.03, F=0.82, p=0.45). The post-interview baseline heart rates were non-significantly almost equal in the traditional, ChatGPT voice app, and mixed reality AI groups, mean (SD)= 83.25 (17.31), 84.84 (18.21), and 82.32 (10.81), respectively (Partial η^2=0.005, F=0.13, p=0.88). There was no statistically significant difference in the baseline heart rate between the pre-interview and post-interview in any of the groups (p=>0.05) (Table 27). Figure 14 also shows slight differences between groups in the baseline heart rate in the pre- and post-tests.

Table 27

Difference between study groups in mean baseline heart rate pre-and post-interview (beats per minute) (N=55)

Heart rate at baseline		**Traditional (n=15)**	**ChatGPT voice app (n=20)**	**Mixed reality AI (n=20)**	**Partial η^2**	**F**	***p* value** [a]
Pre-interview	**Mean (SD)**	93.89 (16.25)	87.26 (16.09)	88.86 (14.39)	0.03	0.82	0.45
	Median (IQR)	97.45 (81.86, 107.88)	84.46 (75.14, 102.41)	88.53 (77.44, 101.91)			
	Min-Max	63.40 – 114.75	60.56 – 116.33	60.56 – 107.23			

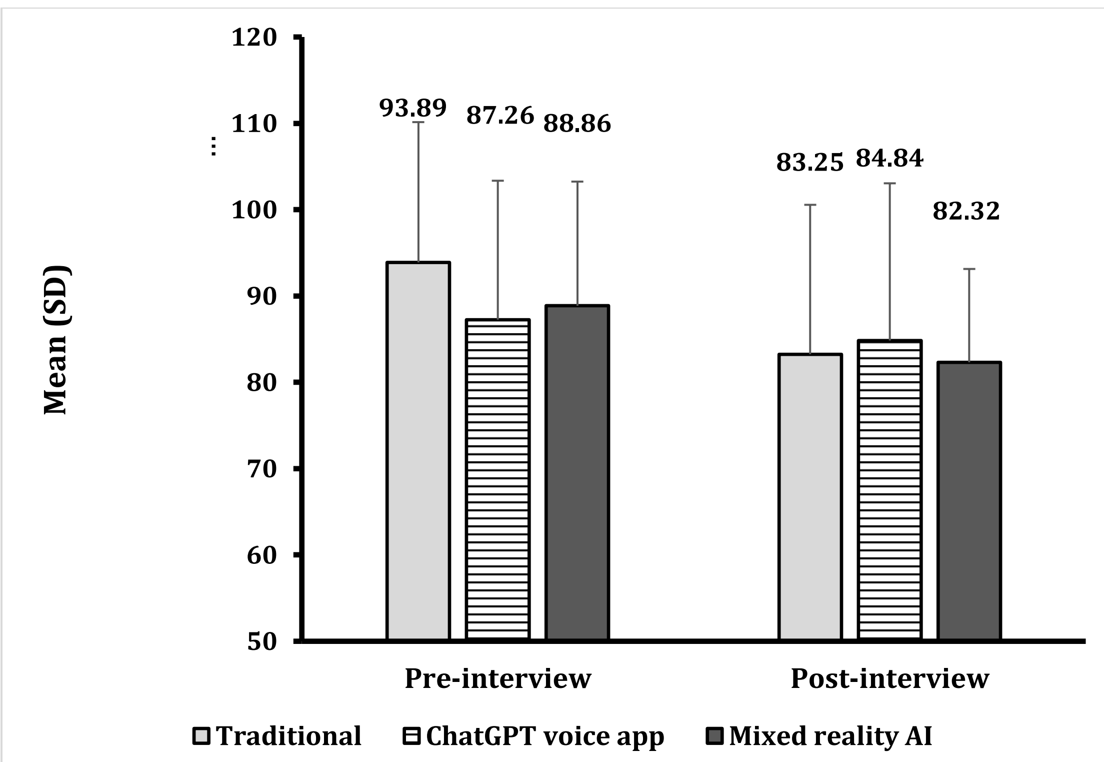
Post-interview	**Mean (SD)**	83.25 (17.31)	84.84 (18.21)	82.32 (10.81)	0.005	0.13	0.88
	Median (IQR)	79.66 (70.63, 97.46)	86.97 (72.82, 96.25)	82.37 (72.48, 89.97)			
	Min-Max	57.96 – 116.43	51.95 – 117.40	64.63 – 102.00			
Mean difference (SD)		10.65 (22.83)	2.43 (25.11)	6.54 (17.25)	-		
t		1.81	0.43	1.69			
95% CI		(-2.00, 23.29)	(-9.33, 14.18)	(-1.54, 14.61)			
***p* value** [b]		0.09	0.67	0.11			

SD Standard deviation, IQR Interquartile range, Min Minimum, Max Maximum, CI Confidence interval, η^2 Eta squared

[a]One-way ANOVA, [b]Paired t-test

Figure 14

Difference between study groups in mean baseline heart rate pre-and post-interview



In repeated measures of ANOVA, there was no significant association between post-interview baseline heart rate and the study groups (Partial η^2 = 0.01, p=0.71). The post-interview baseline heart rate was not significantly associated with the interaction between the study groups and the pre-interview baseline heart rate (Partial $\eta^2 = 0.02$, p=0.55) (Table 28).

Table 28

Association between post-interview baseline heart rate, study groups, and the interaction between study groups and pre-interview baseline heart rate (beats per minute) (N=55)

Variables	**F**	**Partial η^2**	***p* value**
Group	0.34	0.01	0.71
Group * Pre-interview heart rate	0.61	0.02	0.55

η^2 Eta squared
Repeated measures of ANOVA test

In linear regression, the ChatGPT voice app group was not significantly associated with post-interview baseline heart rate (B=15.92, 95%CI: -43.89, 75.73, p=0.60). The mixed reality group was not significantly associated with the post-interview baseline heart rate (B=1.05, 95%CI: -62.07, 64.16, p=0.97). Pre-interview baseline heart rate was not significantly associated with post-interview baseline heart rate (B=0.08, 95%CI: -0.41, 0.57, p=0.75). The interactions between the ChatGPT voice app and pre-interview baseline heart rate, as well as between mixed reality AI and pre-interview baseline heart rate, were not significantly associated with post-interview baseline heart rate (B= -0.16, 95%CI: -0.81, 0.49, p=0.63) and (B= -0.02, 95%CI: -0.70, 0.66, p=0.96). The model explained 9% of the variance in the post-interview baseline heart rate (Table 29).

Table 29

Linear Regression of the Association Between Post-Interview Baseline Heart Rate, Study Groups, Pre-Interview Baseline Heart Rate, And the Interaction Between Study Groups And Pre-Interview Baseline Heart Rate (Beats Per Minute) (N=55)

Explanatory variables		**B (95% CI)**	**SE**	***p* value**
Group	**Traditional**	Reference category		
	ChatGPT voice app	15.92 (-43.89, 75.73)	30.52	0.60

	Mixed reality AI	1.05 (-62.07, 64.16)	32.20	0.97
Pre-interview heart rate at baseline		0.08 (-0.41, 0.57)	0.25	0.75
Traditional * Pre-intervention score		Reference category		
ChatGPT voice app * Pre-intervention score		-0.16 (-0.81, 0.49)	0.33	0.63
Mixed reality AI * Pre-intervention score		-0.02 (-0.70, 0.66)	0.35	0.96

Adjusted R squared= 0.091
B Regression coefficient, SE Standard error

The pre-interview during-interview heart rate was non-significantly associated with the study groups, where the mean (SD) pre-interview during-interview heart rates in the traditional, ChatGPT, and mixed reality AI groups were 91.62 (15.42), 84.53 (9.85), and 84.38 (10.90), respectively (Partial η^2=0.07, F=1.96, *p*=0.15). The post-interview during-interview heart rate was non-significantly associated with the study groups, where the mean (SD) post-interview during interview heart rates in the traditional, ChatGPT, and mixed reality AI groups were 82.82 (17.65), 86.82 (19.20), and 81.96 (13.06), respectively (Partial η^2=0.02, F=0.47, p=0.63). There was no statistically significant difference in the pre-interview and post-interview heart rates among the groups (*p*=>0.05) (Table 30). Figure 15 shows the difference between the study groups in the mean during-interview heart rate and pre- and post-interview HR.

Table 30

Difference Between Study Groups In Mean During-Interview Heart Rate Pre-And Post-Interview (Beats Per Minute) (N=55)

Heart rate during interview		**Traditional (n=15)**	**ChatGPT voice app (n=20)**	**Mixed reality AI (n=20)**	**Partial η^2**	**F**	***p* value** [a]
Pre-interview	**Mean (SD)**	91.62 (15.42)	84.53 (9.85)	84.38 (10.90)	0.07	1.96	0.15
	Median (IQR)	89.87 (84.21, 101.97)	85.84 (74.86, 93.39)	86.34 (75.90, 91.45)			
	Min-Max	62.89 – 117.58	65.02 – 98.45	60.56 – 98.80)			

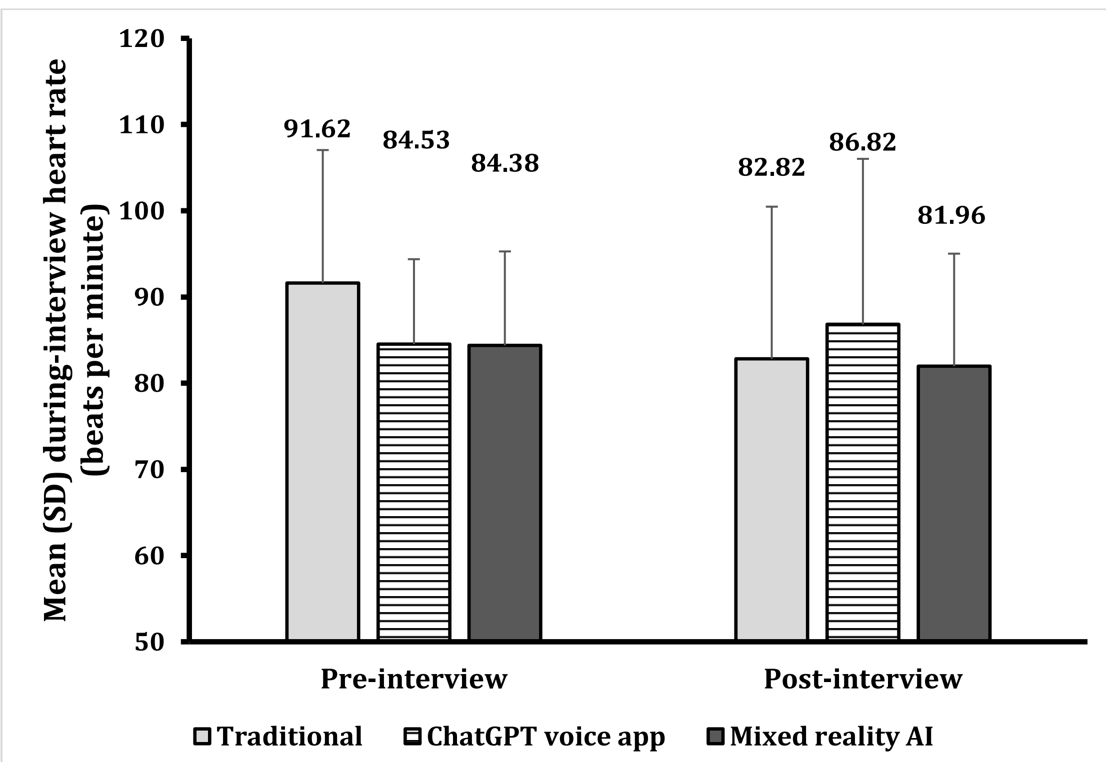
Post-interview	**Mean (SD)**	82.82 (17.65)	86.82 (19.20)	81.96 (13.06)	0.02	0.47	0.63
	Median (IQR)	80.61 (68.00, 98.35)	90.95 (77.53, 97.90)	81.48 (72.37, 93.95)			
	Min-Max	50.50 – 109.39	43.71 – 116.96	55.91 – 102.71			
Mean difference (SD)		8.80 (24.78)	-2.29 (19.74)	2.42 (18.69)	-		
t		1.38	-0.52	0.58			
95% CI		(-4.92, 22.53)	(-11.53, 6.95)	(-6.33, 11.17)			
***p* value** [b]		0.19	0.61	0.57			

SD Standard deviation, IQR Interquartile range, Min Minimum, Max Maximum, CI Confidence interval, η^2 Eta squared

[a]One-way ANOVA, [b]Paired t-test

Figure 15

Difference between study groups in mean during-interview heart rate pre-and post-interview



In repeated measures of ANOVA, there was no statistically significant association between post-interview during-interview heart rate and study groups (Partial η^2 = 0.05, p= 0.26). The interaction between study groups and pre-interview heart rate during the interview was not significantly associated with post-interview heart rate during the interview (Partial η^2 = 0.04, p= 0.31) (Table 31).

Table 31

Association Between Post-Interview During-Interview Heart Rate, Study Groups, And The Interaction Between Study Groups And The Pre-Interview During-Interview Heart Rate (Beats Per Minute) (N=55)

Variables	F	Partial η^2	*p* value
Group	0.72	0.03	0.49
Group * Pre-interview score	1.21	0.04	0.31

η^2 Eta squared
Repeated measures of ANOVA

In linear regression, the ChatGPT voice app group was not significantly associated with post-interview heart rate during the interview (B=-41.54, 95%CI: -121.70, 38.61, *p*=0.31). The mixed Reality group was not significantly associated with post-interview during-interview heart rate (B=7.98, 95%CI: -67.60, 83.55, *p*=0.84). Pre-interview during-interview heart rate was non-significantly associated with post-interview during-interview heart rate (B=-0.14, 95%CI: -0.68, 0.41, p=0.62). The interaction between the ChatGPT voice app and pre-interview heart rate was not significantly associated with post-interview heart rate (B=0.53, 95%CI: -0.38, 1.44, p=0.26). The interaction between mixed reality AI and pre-interview during-interview heart rate was not significantly associated with post-interview during-interview heart ratepre-and12, 95%CI: -0.97, 0.74, p=0.79) (Table 32).

Table 32

Linear Regression of the Association Between Post-Interview During-Interview Heart Rate, Study Groups, Pre-Interview During-Interview Heart Rate, and The Interaction Between Study Groups and the Pre-Interview During-Interview Hear Rate (Beats Per Minute

Explanatory variables	B (95% CI)	SE	*p* value

Group	**Traditional**	Reference category		
	ChatGPT voice app	-41.54 (-121.70, 38.61)	40.90	0.31
	Mixed reality AI	7.98 (-67.60, 83.55)	38.56	0.84
Pre-interview scores		-0.14 (-0.68, 0.41)	0.28	0.62
Traditional * Pre-intervention score		Reference category		
ChatGPT voice app * Pre-intervention score		0.53 (-0.38, 1.44)	0.47	0.26
Mixed reality AI * Pre-intervention score		-0.12 (-0.97, 0.74)	0.44	0.79

Adjusted R squared = -0.046
B Regression coefficient, SE Standard error

The pre-intervention FLCAS was non-significantly higher in the traditional than in the ChatGPT voice app or mixed reality AI groups, mean (SD) = 94.47 (12.82), 86.35 (19.91), and 88.20 (16.02), respectively, (Partial η^2=0.04, F=1.06, p=0.36). The post-intervention FLCAS was non-significantly lower in the mixed reality AI group than in the traditional and ChatGPT voice app groups, mean (SD)= 83.75 (17.84), 94.40 (11.08), and 90.40 (24.34), respectively, (Partial η^2=0.05, F=1.40, p=0.26). There was no statistically significant difference in the pre- and post-intervention FLCAS scores in any group (p=>0.05) (Table 33). Figure 16 shows the difference between the study groups in the mean FLCAS pre- and post-intervention.

Table 33

Difference Between Study Groups in Mean Foreign Language Classroom Anxiety Scale Pre-And Post-Intervention (N=55)

FLCAS		**Traditional (n=15)**	**ChatGPT voice app (n=20)**	**Mixed reality AI (n=20)**	**Partial η^2**	**F**	***p* value [a]**
Pre-intervention	**Mean (SD)**	94.47 (12.82)	86.35 (19.91)	88.20 (16.02)	0.04	1.06	0.36
	Median (IQR)	97.00 (84.00, 100.00)	84.00 (77.00, 96.00)	93.00 (75.00, 97.75)			
	Min-Max	74.00 – 116.00	61.00 – 145.00	45.00 – 116.00			
	Mean (SD)	94.40 (11.08)	90.40 (24.34)	83.75 (17.84)	0.05	1.40	0.26

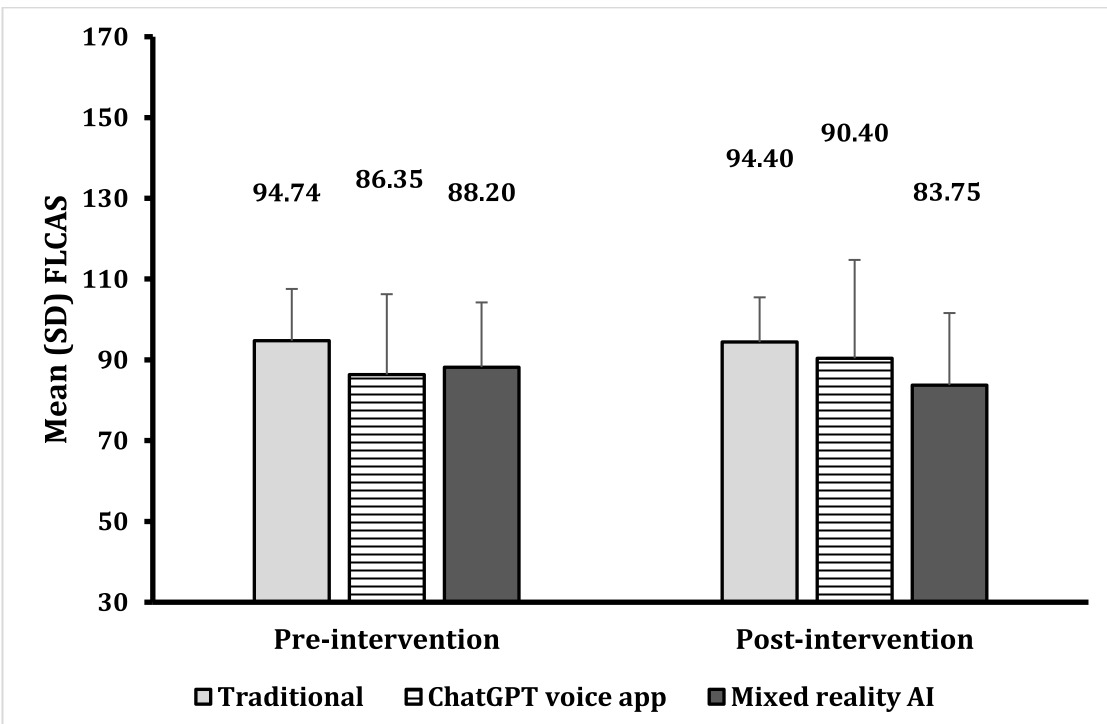
Post-intervention	**Median (IQR)**	92.00 (85.00, 104.00)	91.00 (80.75,102.75)	88.00 (65.50, 96.25)			
	Min-Max	81.00 – 104.00	46.00 – 147.00	56.00 – 117.00			
Mean difference (SD)		0.07 (10.00)	-4.05 (21.58)	4.45 (22.08)	-		
t-statistics		0.03	-0.84	0.90			
95% CI		(-5.47, 5.61)	(-14.15, 6.05)	(-5.88, 14.78)			
***p* value** [b]		0.98	0.41	0.38			

SD Standard deviation, IQR Interquartile range, Min Minimum, Max Maximum, CI Confidence interval, η^2 Eta squared

[a]One-way ANOVA, [b]Paired t-test

Figure 16

Difference between study groups in mean Foreign Language Classroom Anxiety Scale pre-and post-intervention



In repeated measures of ANOVA, there was no significant association between post-intervention FLCAS and study groups (Partial $\eta^2 = 0.05$, $p = 0.26$). The interaction between the study groups and pre-intervention FLCAS was not significantly associated with post-intervention FLCAS (Partial $\eta^2 = 0.04$, p=0.39) (Table 34).

Table 34

Association between post-intervention total Language Classroom Anxiety Scale, study groups, and the interaction between study groups and pre-intervention scores (N=55)

Variables	F	Partial η^2	*p* value
Groups	1.37	0.05	0.26
Groups * Pre-intervention FLCAS	0.96	0.04	0.39

FLCAS Language Classroom Anxiety Scale, η^2 Eta squared
Repeated measures of ANOVA

In linear regression, the ChatGPT voice app group was not significantly associated with post-intervention FLCAS (B=-7.25, 95%CI: -79.18, 64.69, *p*=0.84). The mixed reality AI group was not significantly associated with post-intervention FLCAS (B=28.08, 95%CI: -48.10, 104.26, *p*=0.47). Pre-intervention FLCAS was non-significantly associated with post-intervention FLCAS (B=0.57, 95%CI: -0.10, 1.24, *p*=0.10). The interaction between ChatGPT voice app and pre-intervention FLCAS was non-significantly associated with post-intervention FLCAS (B=0.09, 95%CI: -0.68: 0.86, *p*=0.82). The interaction between mixed reality AI and pre-intervention FLCAS was not significantly associated with post-intervention FLCAS (B= -0.40, 95%CI: -1.21, 0.42, *p*=0.34) (Table 35).

Table 35

Linear Regression of The Association Between Post-Intervention Total Language Classroom Anxiety Scale, Study Groups, Pre-Intervention Score, And the Interaction Between Study Groups And Pre-Intervention Score (N=55)

Explanatory variables		B (95% CI)	SE	*p* value
Groups	**Traditional**	Reference category		
	ChatGPT voice app	-7.25 (-79.18, 64.69)	36.70	0.84

	Mixed reality AI	28.08 (-48.10, 104.26)	38.87	0.47
Pre-intervention FLCAS		0.57 (-0.10, 1.24)	0.34	0.10
Traditional * pre-intervention FLCAS		Reference category		
ChatGPT voice app * pre-intervention FLCAS		0.09 (-0.68, 0.86)	0.39	0.82
Mixed reality AI * pre- intervention FLCAS		-0.40 (-1.21, 0.42)	0.42	0.34

R squared = 0.52

FLCAS Language Classroom Anxiety Scale, B Regression coefficient, SE Standard error

Analysis of Research Question 3 (the NoM strategies)

This section identifies the negotiation of meaning strategies exhibited by the two experimental groups: the ChatGPT audio group and the MECAI in MR group. Over six weeks, participants engaged with the AI tools they were assigned to for 20 minutes, which was part of a 4-hour class activity. The analysis focused on identifying patterns of negotiation of meaning that emerged when both experimental groups interacted with two different AI applications to improve their speaking and comprehension of job interview questions.

After six weeks of practice with both AI tools, the ChatGPT audio app and the MECAI in MR, the conversation logs of each group were collected and examined. Although some students in the ChatGPT group faced technical challenges (e.g., lost conversational logs and connectivity issues), all participants completed the six-week practice. One concern is that students in the ChatGPT group may have taken some responses for granted, engaging in limited negotiation of meaning, such as clarification requests, confirmation checks, or other interactional strategies. Conversely, it was much easier to retrieve the conversation logs for the MR groups because they were connected to a single backend. There were minor challenges in connecting the headsets to the Internet, but they were resolved on the first day. Room scanning with the HMD to identify furniture for accurate placement of the virtual environment also took some time during the initial session, but in subsequent weeks the headset automatically recognized the furniture and physical environment, so it did not take much time in later practice sessions.

After collecting the conversation logs, 95,536 words were collected from all participants in the ChatGPT group, whereas the MR group produced 146,968 words. Normalization between both corpora was calculated to provide percentage of NoM per week. Students using MECAI in mixed reality (MR) requested additional time to practice, but to ensure fairness, the researcher asked them to use the MECAI in MR app during their allocated time. Two of these students, who had upcoming job interviews, reported that the MECAI in MR was particularly useful for their interview preparation. Although they were already familiar with ChatGPT, they perceived the MR-based experience as more beneficial.

Tables 36 and 37 show the NoM strategies percentage per 1000 words per condition. It is clear that the MECAI in the MR group used more NoM strategies than the ChatGPT group.

Table 36

MECAI In MR Group NoM Strategies Percentage Per 1000 Words After Normalization (Sessions 1–6)

NOM Function	S1	S2	S3	S4	S5	S6
Elaboration	36.73%	43.16%	43.10%	43.28%	51.22%	58.33%
Clarification Request	27.55%	24.21%	27.59%	20.90%	23.17%	19.44%
Confirmation	14.29%	9.47%	13.79%	11.94%	10.98%	9.72%
Confirmation Check	11.22%	14.74%	7.76%	14.93%	7.32%	6.94%
Elaboration Request	5.10%	4.21%	4.31%	4.48%	3.66%	2.78%
Correction/Self-Correction	4.08%	3.16%	2.59%	2.99%	2.44%	1.39%
Vocabulary Check/Request	1.02%	1.05%	0.86%	1.49%	1.22%	0%
Reply Clarification/Def.	0%	0%	0%	0%	0%	1.39%
SESSION TOTAL	100%	100%	100%	100%	100%	100%

Table 37

ChatGPT Voice Group NoM Strategies Frequency (Sessions 1–6)

NOM Function	S1	S2	S3	S4	S5	S6
Elaboration	49.61%	57.26%	55.67%	55.97%	50.00%	49.57%
Clarification Request	24.03%	15.38%	22.16%	11.94%	14.66%	17.95%
Confirmation	15.50%	14.53%	15.46%	11.19%	13.79%	11.11%
Confirmation Check	1.55%	1.71%	1.03%	0.75%	0%	0%
Elaboration Request	8.53%	10.26%	5.15%	20.15%	21.55%	20.51%
Correction/Self-Correction	0.78%	0.85%	0.52%	0%	0%	0%
Vocabulary Check/Request	0%	0%	0%	0%	0%	0%
Reply Clarification/Def.	0%	0%	0%	0%	0%	0.85%
SESSION TOTAL	100%	100%	100%	100%	100%	100%

For the traditional group, the researcher observed the students' reactions to instructions during traditional classroom activities. In some sessions, the researcher served as the instructor, while in others, another instructor led the class, and the observations were recorded in writing. Students in the traditional group were usually silent, and whenever they had difficulty understanding the instructor, they asked their peers in their native language a few times when they should have responded to their instructor's questions. This was strikingly different from the AI groups conversation logs who expressed to the researcher that they do not want to use their native language with AI when some students who used Android phones found that ChatGPT responses were in Spanish by default. Students' conversation logs did not include any language other than English. The AI in the MR application only included the English model in the design so that if any student wanted to take advantage, the AI response would only be in English.

Analysis of Research Question 4: Reflection Interviews

This section presents the thematic analysis of the structured interviews with participants from the three conditions. The students were asked 11 questions about their experiences with their conditions, the role of the instructor and pedagogical framework, and how it improved their speaking performance and reduced their anxiety. The fourth research question examines the students' experience in the class as a whole and how their conditions filled gaps in terms of teachers' knowledge and curriculum gaps. This analysis was guided by the previous literature to answer the question: What are the perceptions of the students in the experimental and control groups about the effectiveness of these conditions in helping them overcome their anxiety and improve their speaking performance?

The qualitative data revealed four overarching themes that showed the participants' perceptions of their learning experiences under each condition:

Figure 17 *Word Cloud From NVIVO 15 Demonstrating the Most Prevalent Themes*

According to the Word Cloud extracted from NVIVO 15 (Figure 17), the most prevalent themes underscored the value of practicing for job interviews and helped improve their knowledge. The word ChatGPT was a frequent theme in the students' responses, as was the word person, which was considered a common word among the ChatGPT group and the MECAI in the MR group. Both groups considered AI as a real person, an old friend, or a brother in the case of ChatGPT. Therefore, a detailed account of all qualitative data results is described below.

Theme 1. The Role of AI In Reducing Students' Anxiety And Improving Their Speaking Performance

This theme included different subthemes and sub-subthemes. All students felt comfortable using both conversational AI agents, ChatGPT and the conversational AI agent in MR. However, the two groups' perceptions differed in terms of how they viewed practicing with AI. The themes were divided into four subthemes:

Theme 1: Role of AI in Reducing Anxiety & Improving Speaking

1.1 First impression about Treatment

1.2 ChatGPT & Anxiety

1.3 ChatGPT & Speaking Performance

1.4 MECAI in MR & Anxiety

1.5 MECAI in MR & Speaking Performance

1.5.1 Realism and Social Presence

1.6 Condition Preference

1.1. First impression about treatment

The themes related to the students' experiences with both treatments, the conversational AI agent in MR and the ChatGPT voice application. Students' first impressions of both technologies ranged from positive impressions of using AI in general to curiosity and uncertainty at certain times.

The first subtheme relates to students' impressions of AI when it was introduced and explained to them. The question attempts to capture their first impression and then link it to their final impressions in terms of recommending their condition or not. Students who used conversational AI agents in MR had a positive impression. For example, one student described their experience as follows:

"My first impression when I put the mixed reality. That was the first time that I used this device. For me that's very good impression."

Some students described challenges with the way they communicate with ChatGPT at their initial experience, but finally they were proud of their performance. For example, one student described their teacher as follows: "My teacher showed me how to use ChatGPT to practice job interviews. I felt a little nervous at first. I didn't know how it worked. But it was easy and interesting. I liked that I could practice alone. After trying it, I felt better. It was a good way to learn."

The experience at the very beginning was challenging for the students using the ChatGPT voice app or the MECAI group in MR, but they mentioned that it was rewarding in the end. The reasons for this rewarding experience are explained under the following themes. Ease of use was also a recurring theme, and teachers' help in understanding how to use AI throughout the six weeks of practice was a common theme in the reflection interview.

Nervousness was a common sub-theme for the condition's first impression. This is due to two main reasons: low proficiency levels and lack of experience with conversational AI. Students described their feelings as "curious," "nervous," "lost," "upset," and "scared" due to uncertainty about what AI would ask them and how they would respond.

1.2 ChatGPT & Anxiety

Students' reflections on this theme can be divided into three components: personalization, anxiety-free environment, and lack of embodiment.

In the students' reflections, they mentioned how they did not know how to write a resume or prepare for a job interview, and ChatGPT taught them how to perform these tasks professionally, which reduced their anxiety.

Creating an environment that is personalized and free of judgement was also a recurring theme in the interviews. Students referred to the benefits of practicing with AI as it allowed them to practice individually and created an environment free of judgement:

"Because I was able to practice often and without Judgement."

The students also reflected on classroom teaching and incorporating ChatGPT into their activities.

"I feel better in ChatGPT Voice App and when you teach me"

For the students, practicing job interview questions with ChatGPT was helpful, as it made them relax and reduced their anxiety.

"Practicing helped. I relaxed. I don't stumble as much anymore. It was helpful."

Being impressed by AI functionality and how it can help students practice and know more about jobs that the teacher might lack knowledge about.

"I really impress with it because it tells me what I want to hear."

The way AI was empathetic with the learner, giving them positive feedback and correcting them at times in "a soft way" had a positive impact on a student.

"It was impact me that the AI Congrats me and correct me in a soft way."

Students also saw ChatGPT as a brother or an AI companion that reduced their anxiety.

"When I'm talking to ChatGPT, I imagine I'm talking to my brother."

However, students also reflected on ChatGPT's disembodiment and lack of human connection. Students described ChatGPT, for example as

"No face," "I couldn't see facial expressions," but a student described these as "small problems compared to the benefit, " and another student stated that "even they no have face but it's helpful."

This theme encompassed several ideas that students reflected upon. Students reflected on ChatGPT as an object-to-think-with:

"Finally, made checks and all and all. When the AI appear and I discover that I have the opportunity to save a lot of time and teach me also that the AI cannot do the work of a human. It needs to be used as a tool to something"

Students' experiences in practicing job interviews with ChatGPT were based on its perceived usefulness. They can practice several times, and the AI tool can personalize its responses, helping students become aware of their mistakes. It also helped them prepare for real job interviews as it simulated real-life situations and "realistic questions." Students also commented on how the teacher integrated AI to complement classroom materials and provided tips for students.

"The ChatGPT Voice App was very effective in helping me improve my speaking."

"ChatGPT taught me what is a resume, how to answer the questions and how to prepare for interview and gave more ideas and opinions."

"I feel very thankful to show me different example. The are different ways to prepare for an interview and I think that you give us a lot of these example. In addition, the common and personal tips are nice, and I take note of a majority of these. It is important to know how to do it an interview in paper, and you give us different tips that the class materials do not."

"I liked that I could speak freely and get helpful feedback. The app didn't judge me, so I wasn't afraid to make mistakes. I could listen to my answers and notice what I needed to change."

"This helped me become more aware of my grammar, vocabulary, and pronunciation."

However, students shared that ChatGPT can be challenging when it comes to missing important non-verbal feedback, body language, facial expressions, and emotional responses that can guide feedback. Students also reflected on the teaching method that incorporated AI for 20 minutes within four hours of class. This method of teaching encouraged students to learn and reduced their anxiety.

"I feel better in ChatGPT Voice App and when you teach me,"

"One challenge I faced was the lack of non-verbal feedback from the ChatGPT Voice App. In a real interview, body language and facial expressions matter a lot, and I didn't get any of that here."

Practicing with AI gave students more confidence in performing real job interviews. One student shared that practicing with ChatGPT made them feel confident in a real job interview.

"I recently had a real interview in English and felt much more calm and prepared."

1.3 ChatGPT & Speaking Performance

For this sub-theme, when students were asked if their speaking performance improved after using ChatGPT, they linked their feelings of comfort to their speaking performance improvement. Lack of human engagement at times when the students were using ChatGPT made them feel less anxious when "no one was watching." This made the students focus on improving their skills, and the more they practiced, the better they became.

"But with the app, I could practice in a private and comfortable way. There was no one watching me, so I could relax and focus on improving. I noticed that the more I practiced, the less anxious."

“yes, because when I asked ChatGPT it give me properly answers so now I know how to properly answer an interview.”

ChatGPT was also a game changer for some of the students. It changed their perspective when it came to knowing how to answer job interview questions. The response below links comfort with improved practice:

“One thing I would like to share is how much this practice changed the way I feel about speaking English”

When asked about their perception of the usefulness of ChatGPT, a student responded positively about the teaching method in general, which integrated AI.

“I think is good in the way we learn.”

1.4 MECAI in MR & Anxiety

This subtheme revolved around how the conversational AI agent in MR helped students improve their speaking skills.

Most students in the MR condition considered the conversational AI agent as a “person,” “someone,” or an “older friend.” They reflected on how the AI agent was effective in improving their understanding, giving them recommendations, making them relax when practicing

“She was effective for me that I can understand, and I can express what I can do.”

“When I began to talk with her, I said, I am going to give you some questions, and then you ask me these questions". She asked me question by question and when I finished to answer she gave me recommendations for do this better.”

“When it’s a person I have more anxiety, but AI made me feel less anxiety. Right now, I feel good if I will have an interview with a person. Because I feel prepared.”

Other students expressed challenges in terms of not being able to get their questions across because of their accent. Some students initially did not understand the vocabulary words the MECAI in MR used, and it was difficult for them, especially when ChatGPT was so fast in responding. However, they gradually became able to communicate with AI.

“it doesn’t understand me”

“Sometimes I don't like it so much because he speaks very fast and says many things at the same time and it's confusing for me.”

“Because the question and the answer are in English, and sometimes I understand some vocabulary is hard too.”

For this group, two students were asked to use the Conversational AI agent in MR before real job interviews. One of them asked to stay after class because she had a job interview the following day.

“I like it a lot. I felt real. I asked my teacher to use it more before a real job interview.”

Students also reflected on how their practice with AI improved their performance and their speaking skills.

“Before I say little. Now I say more.”

“Its help me a lot improving my skills”

Many interviewees reflected on the conversational AI agent in the MR for its support. Their responses included the AI’s ability to give advice, engaging responses that impacted attention, model examples, providing interactive conversations, learning gains that impacted how they responded to questions, considering embodied AI as a good opportunity to practice, helping with pronunciation, and improving speaking.

This can be seen in the following responses:

“Of course. Yes, because she gave me a lot of advice.”

“So, the most important thing for me is that I paid attention with AI.”

“And she give me, like many example about that question.”

“have more knowledge about how to answer this question.”

1.5 MECAI in MR & Speaking Performance

Students' reflections on this part helped us understand how AI also improved their speaking performance.

One student asked the MECAI in MR how to become a nurse as part of their preparation for a job interview in this field, and the AI provided detailed information about programs and degrees.

“When I asked her how I can become a nurse she give many methods that made me have more self-confidence”

A student also had a real job interview and expressed that the use of AI made this experience less stressful.

“I had a job interview, and I was veery scarred but with AI experience about the interview how to attend the interview and what should I do.”

“I felt nervous with my skills speaking but during my practices I could without nervous. In this moment I feel complete confidence with my abilities in a job interview.”

Several students described AI as “helpful,” made them “feel safe,” “helped me with my confidence,” and provided positive feedback.

This can be explained by the following excerpts from student reflections:

“The moment that most impacts my confidence was when I finished to answer the questions and she said, "well done". When she said that I got most confidence with my skills speaking.”

“she [instructor] made less worry.”

Students in this condition also acknowledged the use of AI to complement human knowledge and the opportunity to practice. This will also be laid in the theme of the teacher’s role.

“We are human, and our knowledge are like limited.”

For students, MECAI in MR mimics authentic situations that look real and can be transferred to real situations. Students also reflected on its effectiveness due to “comprehensive and expert responses.” Some students stated that AI has more specialized knowledge than the Internet.

Students believed that MECAI in MR helped them organize their responses and improved their listening and pronunciation skills, which improved their speaking performance in comparison to the pretest mock job interview. The MECAI in MR improved their performance as they could express their feelings and concerns. They also reflected on its advantages in terms of personalization and their ability to practice individually and recommended their condition as AI was more knowledgeable.

However, some students noted repeated response patterns in AI that could be challenging and cause frustration. In addition to language comprehension challenges, AI may not fully understand the students or vice versa. However, students mentioned that this could be resolved by asking the AI to repeat or explain.

"Yes... because the question and the answer are in English, and sometimes I understand some vocabulary is hard to... understand the meaning. In that case, when you don't have clear... the words... have the the idea."

1.5.1 Realism and Social Presence

This sub-sub-theme relates to the speaking performance as the students showed that communicating with MECAI in MR felt real like a job interview. The way the AI agent was designed in an office reflected a real job interview experience, made the students feel presence in the MR environment.

Many students in this MR condition described the AI agent as a "real person," "woman," or an "older friend" or "someone" who gave advice and instructions. Some participants who had experience with ChatGPT prior to the study believed that interacting with an embodied AI agent was like interacting with a human compared to ChatGPT, which does not have a face. They also commented on the AI agent "following them with her eyes" (i.e., referring to eye-tracking). They also mentioned that the way the office was designed felt like a real interview.

However, two students expressed different opinions about the AI agent environment in MR.; one said that she paid more attention to AI than she did in class. She felt that it was more personalized. Another student mentioned that she was distracted by the bottles and books placed on the desk. The environment was deliberately and purposefully designed by the researcher to avoid grabbing objects or hand interaction.

This can be shown in the following excerpts from students' responses in the AI agent MR group:

"The AI talked like real person."

"That experience is different because I feel like I am talking with somebody I am asking question, and she gives me answer, and I can see that person. It's different. Because ChatGPT or another AI tool... AI, you just write and you get answer."

1.6 Condition Preference

To understand whether the three groups preferred another condition or the condition they were placed in, all students were asked in the interview about their opinion regarding their condition and whether they wanted to recommend it to someone else.

Control Group

When asking the control group whether they preferred their condition or the other conditions, their responses reflected how teacher-led instruction provided the students with feedback tailored to their needs. Many students found this aspect to be highly beneficial. This was demonstrated in their opinion about how a human instructor can have a depth of knowledge that can help learning. The students reflected on teacher-led instruction as an opportunity to explore not only job interviews but also the cultural nuances of job interview situations. For them, teacher-led instruction created a safe, adaptive environment, and the students felt supported, less pressured, and able to make mistakes without fear of judgement.

This interaction with a human tutor allows for personalized communication and immediate feedback. The human connection was a concept that many students referred to, as the teacher can notice the students body language and facial expressions and adjust their language accordingly to respond to the students' emotions in real time.

ChatGPT Group

Students in the ChatGPT condition expressed their preference for using ChatGPT as it was less intimidating and provided a more flexible and accessible way to practice English. Its ubiquity and presence on different platforms allow for ease of use and practice at any time. Students appreciated the way AI provided personalized learning that reduced their anxiety and gave them control over their practice.

They also expressed their preference for ChatGPT as it helped them with vocabulary and pronunciation. They also felt less judged and were more willing to experiment with language. It was also observed that when they interacted with ChatGPT, they mentioned that they did not want to speak to it using their mother tongue. This was strikingly interesting as the students in the traditional condition used their mother tongue to ask their peers questions when they did not understand the teacher. However, some students noted that the lack of human presence during practice reduced opportunities for human interaction that would teach them body language skills and non-verbal communication.

MECAI in MR Group

The MR group expressed a strong preference for their condition due to realism and an immersive environment that made students feel as if they were in a real job interview. Students reflected on their immersive experience, as MR made the interview practice feel authentic and engaging. This feeling of social presence and situated learning helped them prepare for real-life job interviews. The sensory-rich immersive environment provided a deeper level of embodiment that neither a traditional classroom nor a ChatGPT application could replicate.

Students in the AI agent MR condition said that MR made them see, hear, and feel as if they were in an actual job interview. This immersive environment helped them concentrate on using language. For some, MR was less intimidating than face-to-face practice and more engaging than practice with ChatGPT. The table below lists some references for these themes.

2. Teacher Role in AI-Mediated Language Learning

In their reflection interviews, the students reflected on different aspects of the teacher's role. Qualitative data show that students deeply value human teachers as guides, motivators, and sources of confidence. They consistently placed stress on different aspects of the teacher role and teaching method, whether it targeted ChatGPT and AI agents in MR in their learning in the treatment groups or not in the control group. All groups acknowledged awareness of general human and curriculum limitations, yet they valued the teaching method and appreciated the human element of teaching as an anchor for their learning process.

To evaluate whether the students based their reflections on facts, not bias, or because the researcher was also the teacher, the students were asked to bring examples to show why they thought teaching was effective.

The students in the treatment groups provided examples that indicated how the teacher introduced them to ChatGPT or the AI agent in MR and how the teacher explained how to use each application and the purpose of using them to supplement the curriculum and teacher knowledge.

Students provided examples such as the following:

“I think my teacher can give me ideas but not like ChatGPT. she doesn't have the same knowledge”

A student from the MR group stated:

“We are human, and our knowledge are like limited.”

Students also referred to the fact that AI in general provided personalized learning experiences, unlike the teacher who can be busy with other students

“She [the teacher] might be busy with another student.”

Students also compared the use of AI with the unique role of teachers in terms of offering comfort, encouragement, and emotional presence. What mattered to the students was teacher guidance, correction, and teaching method. Besides the use of AI, students also expressed that they gained confidence from student-centered group work and pair work activities as part of the TBLT teaching method, which had a strong impact on their performance. They repeatedly reflected on constructive encouragement, feeling valued and respected, and the value of traditional teaching methods.

According to the students in the treatment group, integrating ChatGPT within classroom activities was encouraging and improved their speaking skills, but they also pointed out that teacher-led instruction and guidance provided support and encouragement. However, some students from the control group said that a human teacher can increase anxiety due to the fear of judgement and that a teacher corrects every mistake they make. They mentioned patience in onboarding ChatGPT, reduced anxiety through preparation in traditional teaching methods, and that reflection and self-awareness could reduce anxiety among students.

Some examples to illustrate this are as follows:

"But the teacher helped us with the practices in the classroom. I liked to explore the device. As you can see explains very clear. It was great."

"The ChatGPT and your practice help a lot to don't be afraid about an interview. I took the classes as the moment to learn in all my mistakes, and work in these. Help me in hypothetical real interview, practicing with a machine and a classmate. I found a really good improvement anxiety management."

Students also noted that a human teacher can be forgetful or struggle with a bad mood, which might negatively impact them. However, they still believe that a human teacher knows how to teach better than AI because they can also feel the students from their facial expressions and understand students' struggles better than AI. For them, a human teaching method was more trusted than AI. Some students provided examples that illustrated these reflections:

"[teacher} not in a good mood. or forget to let me know what I have to do or respond to a question. because she's a human being."

All groups perceived the role of the teacher as important and principal, while the treatment group also noted the role of AI as supplementary for tasks beyond the teacher's ability and the scope of the curriculum.

Teacher-supportive feedback and encouragement of an environment that enforces the freedom to fail and make mistakes during students' activities was repeatedly mentioned, emphasizing that encouragement and constructive feedback lead to engagement and willingness to communicate, even when students use AI. Immediate feedback helps improve speaking performance, and teacher support builds fluency and confidence through repetition and encouragement. Students mentioned that teacher patience, constructive feedback, and giving students time to respond were crucial for their learning process.

The treatment groups also reflected on the crucial role of the teacher despite their appreciation of AI tools they used. They see AI as a supplementary tool, but not a replacement for the teacher. They acknowledged the fact that AI provides personalization, immediate feedback, and structured practice, but its value is meaningful when guided and integrated by the teacher.

Some examples of these reflections are as follows:

"I see everything you do is good. You do a skill so you help us with grammar, skill, how prepared because this is important for the class you have to do now. Last class we talked with the teacher we learned about skills, but now it's interview. how to prepare for the job interview."

"One moment that helped me a lot was when we did a mock interview in class. You asked me a hard question, "Tell me about the time You solved a problem at work." And at first, I didn't know what to say. But you helped me think of a personal example, and little by little I was able to answer. Everyone in the class listened and supported me. After that, I felt more confident."

"The practice [ChatGPT] helped lower my anxiety about job interviews in English. At first, I felt very nervous when speaking. But you made the class feel safe and calm. You told us it was okay to make mistakes. That helped me relax and try without fear."

All in all, the students reflections suggest that an effective learning environment is not based on teacher only or AI only, but a teaching method that uses sound pedagogy that integrates AI meaningfully, while the teacher role remains central and AI acts as a supplement that provides students with more individualized content to their diverse needs that help them meet the learning outcomes. AI cannot fully detect students' emotions or when they struggle. Students' reflections revealed that AI can fill a gap in the curriculum and teacher knowledge by handling repetitive tasks and providing personalized learning.

3. Anxiety

The final theme in this section is anxiety. This theme was vital to the qualitative findings.. Students reflected on AI "support", ""help with words, "and "specific questions'. Repetition and personalization across both AI tools were recurrent observations among all students. One student mentioned that "I feel more nervous making mistakes in front of a real person." Some students in the control group also noted that they "wished" they had learned how to use AI for job interviews. They mentioned that interacting with a human teacher is helpful, but still makes them "shy" or nervous, like a student from the control group who described his feelings as follows: When you practice with a human, you feel nervous. When they practice with AI, they feel comfortable. "Because it's AI, it's not human." The experimental conditions also mentioned that their first interaction with AI was overwhelming, but the teacher made it easier by showing them how to use it. However, some students across groups noted that anxiety can be healthy and motivating, guiding improvement when tailored instruction is provided. A student mentioned that As my teacher, your feedback was to my level, you explain certain cultural things for certain interview questions."

The human connection was a salient theme across all groups. The integration of AI into the classroom while the teacher has the ability to handle students questions, understand their facial expressions, and their struggle was noted in students' responses. Students mentioned examples such as:

"You could understand when I was struggling or nervous, and you try to help me feel more comfortable."

"I recommend you. I don't use ChatGPT. I recommend you. You teach me how to use the computer. Sometimes I ask you, teacher, how to do this and explain me. Only you talk to us about we can try. We can improve myself with my own writing. you remember, don't use the translator because you want us to try."

It is also worth noting that some students, despite preferring human interaction, still preferred AI for practice purposes. A student from the MECAI in MR group mentioned that "I prefer mix reality because it has more knowledge than teacher". While another student expressed interest in MR for realism and social presence by saying "Yes. Very good. Feel real. I can see, talk. For me is best." Students in the MR group liked the MR experience as it mimics real-life situations some mentioned the the experience felt real "If I used virtual reality, it feel more real." Or they can feel the place, "I like when I see and feel the place." And how the MR environment looked like a real one, "How this looked like a real job interview." A student also expressed shyness when communicating with a human and preferred practice with an AI that has a 'face', "Human good but I feel shy. ChatGPT has no face. I like Mixed Reality. It good for me."

In summary, students across all groups were satisfied with the experience owing to the structural integration of AI into the classroom for a short period. All groups, including the control group, acknowledged the fact that AI can help in terms of repetition, vocabulary, pronunciation, and complementing teachers' knowledge, but human connection remains crucial to the learning process, where teachers can adapt to the learners' needs through their non-verbal communication and can sense their struggle. However, AI can help reduce fear of judgement, help with anxiety, and motivate students through personalized learning.

Chapter V

Discussion

I will discuss the first and second research questions that examined how interaction with conversational AI impacted students' speaking anxiety and performance in comparison with the traditional group. Then I will discuss The third research question investigates the patterns of interaction between the two experimental groups using CMDA. Finally, I will discuss research question 4, which investigated 36 students' perceptions from all three conditions and how they perceived the role of AI conditions, the role of the teacher, the pedagogical framework, and how these factors impacted the students' self-confidence and performance. Emerging themes are also discussed.

Research Question 1

Speaking Improvement Across All Groups with Significant Improvement in MECAI in MR Group

The MECAI in the MR group showed significant improvement in post-mock job interviews, specifically in fluency, vocabulary, grammar, and pronunciation. This aligns with prior research that used situated learning approaches using virtual reality to improve students' speaking (Yan et al., 2024). Moreover, El Shazly (2021) used the Mondly VR app to improve students' speaking performance. Other research shows similar results, noting that it is important to carefully plan the immersive environment to result in speaking improvement (Ironsi, 2023).

Research Question 2

Although there's no statistically significant differences were found in both subjective and objective anxiety measures, this outcome can be best interpreted in terms of the students' pretest scores. Completed pretest FLCAS reports showed low self-reported anxiety and continued to be low at posttest. In contrast, objective heart rate data in the pretest indicated moderate anxiety levels in both the pre- and posttests, with a slight decrease in posttests. This result shows a contradiction between students' subjective assumptions of their anxiety and their actual anxiety during the pre- and posttest mock job interviews. This indicates the limits of self-report instruments in assessing emotional responses, especially in multiethnic populations.

The pretest was challenging for all groups, as shown in their scores. Many of them struggled or skipped questions during the pretest because they lacked task-specific vocabulary, which moderately raised their anxiety during the pretest. Rather than hindering performance, moderate anxiety helped all groups perform better in the posttest. This aligns with research suggesting that moderate anxiety can facilitate performance rather than impede it (Salend, 2011). The educational environment in the classroom at HCCC is supportive and safe, as reported by the students in their reflection interviews. As Toyama and Yamazaki (2021) note, successful learning environments usually maintain a non-threatening atmosphere while allowing a moderate level of arousal to facilitate performance.

This result partially contradicts some studies in which pretests showed high anxiety (e.g., Bao, 2019; Dizon, 2020; Dodigovic, 2007; Tafazoli & Gómez-Parra, 2017; Naseen et al., 2024). Moreover, the slight decrease in anxiety levels in the MR group contrasts with research that found an increase in anxiety in students who used artificial intelligence in immersive environments like Mondly, where heightened cognitive load was identified as a contributing factor (El- Shazly, 2021). In the present study, the MR environment used minimal visual design, did not isolate students from the physical classroom environment, and avoided full 3D immersion. These design affordances of MR likely reduced cognitive load and diminished sensory strain, which may help interpret the absence of high anxiety reported in ElShazly's study (2021), where a conversational AI agent was employed in a virtual reality environment.

Research Question 3

CAI Agent Representation and Modality can Impact NoM Patterns

RQ 3 examined how NoM patterns were employed by the two experimental groups (the ChatGPT voice application and the MECAI in MR groups). Over six weeks of instruction, each weekly class lasted four hours. The teacher integrated the use of AI for 20 minutes using TBLT as a framework, and AI was integrated as part of the main task where students would practice interview questions with AI and ask questions and negotiate meaning to internalize the responses and have good comprehensive input and feedback from AI about their performance. Analysis of the conversational logs indicated that the CAI in the MR group produced more words (146,968) than the ChatGPT voice app group (95,536). Across the six sessions, the CAI in the MR group showed a higher frequency of negotiation strategies (7.77 per 1000 words) than the ChatGPT voice app group (4.12 per 1000 words). Clarification requests and comprehension checks were more frequent in the CAI–MR group than in the ChatGPT voice group. The pattern of NoM and its frequency observed in the CAI in the MR group compared with the ChatGPT voice group indicates that the CAI agent representation and modality (multimodality) can encourage students to communicate and negotiate meaning with AI. Asking the AI to repeat and clarify was more frequent in the CAI group than in the MR group. The CAI in the MR group attempted to overcome communication breakdowns while interacting with AI by asking questions to clarify and confirm understanding. In contrast, the ChatGPT voice group experienced self-paced interactions and took notes during interactions, which led to fewer negotiations of meaning in the later weeks. This discourse behavior underscores how modality can impact the ways students negotiate meaning with AI.

These findings can be interpreted through previous research focusing on synchronous CMC. For example, Pica (1994) describes NoM as the way students modify their input during interaction to overcome comprehension difficulties, whereas Long (1996) views NoM as an opportunity to attend to form and meaning. Therefore, the high frequency of negotiation patterns, such as clarification requests and comprehension checks, in the CAI in the MR group aligns with these theoretical expectations, indicating students' efforts to maintain comprehensible input and modified output. (Gass, 2003). According to Cohn et al. (2021), the more students produce interactions, the more it signifies active learning. This finding also aligns with Van der Zwaard and Bannink (2014), who examined NoM patterns in two different modalities, video and text, and found that the type of modality can result in distinct interaction patterns and frequencies. This study also aligns with research that found that immersive environments can enhance understanding compared to text and voice interactions (Chen & Sevilla-Pavón, 2023). Multimodal interactions, such as using gestures, eye gaze, and facial expressions, can also support higher-level negotiation patterns (Chen & Sevilla-Pavón, 2023; Norris, 2004). This means that embodied interaction in immersive environments such as MR can provide a strong sense of social presence due to the unique affordances of immersive environments, which can facilitate meaningful interactions compared with other computer-mediated interactions (Chen & Sevilla-Pavón, 2023; Norris, 2004). This study also provides new insights, as research on interactions with voice-based and immersive AI agents employing generative Artificial Intelligence is underexplored. This study is also consistent with prior research (Chun, 1994; Sotillo, 2000; Nakahama et al., 2001; Rosell-Aguilar, 2005) that shows that NoM was a crucial mechanism through which learners used NoM strategies to resolve communication breakdowns and improve their understanding of discourse and language skills. This study also aligns with Kim (2017), who found that interacting with voice-based chatbots can provide scaffolding opportunities for students to negotiate meaning and improve their speaking skills.

Research Question 4

Pedagogically informed AI supported by Teacher Guidance Can Improve Learning Outcomes

Research Question 4 examined all participants' perceptions of their experiences in both AI and traditional conditions. As all students were involved in the classroom instructions and utilized different methods of treatment or lack thereof, it was highly important to understand their perceptions of how instruction as a whole and the impact of the three conditions (MECAI in MR, ChatGPT voice, and traditional condition), teaching method, and teacher role impacted students' perceptions regarding reducing their anxiety and improving their speaking performance.

The researcher interviewed 36 of the 55 students. They were asked to answer 11 questions (see Appendix D). The reflection interview included 11 questions about the students' perceptions of their conditions, perceptions of the teacher's role, preference of condition, and whether they could recommend it to others. This section discusses the findings in relation to prior research and underscores the emerging themes.

Role of Conversational AI in reducing Anxiety and improving Speaking Proficiency

The qualitative findings revealed the experimental groups' perceptions of enhanced speaking performance and reduced anxiety, which they attributed to practicing in a judgement-free environment and that AI has more knowledge than the teacher about their diverse interests in different jobs. Students' interaction with conversational AI over six weeks during classroom instruction facilitated language automatization, which was conducive to speaking improvement and less anxiety. This aligns with Ding and Yusuf (2025), and Han and Ryu's (2024) findings that interactions with a conversational AI agent can positively impact students' perceptions and attitudes. Moreover, students had a positive attitude toward AI feedback and how it contributed to their pronunciation improvement and learning. This aligns with Fathi et al. (2024), Ding and Yusuf (2025), and Laksana et al. (2024), underscoring the importance of feedback and its role in overcoming communication obstacles and interlanguage issues (Ding & Yusuf, 2025). The ChatGPT voice group had positive opinions about the accessibility and convenience of using ChatGPT, benefits in improving speaking performance, reduced anxiety, and fluency, which aligned with Laksana et al. (2024) and Cai et al. (2023).

All groups reflected on the importance of human connection in complementing AI interaction. All students reflected on the role of the teacher in discerning students' needs from their facial expressions, voice tone, and attitude in the classroom, which might not be detected by AI. This aligns with myriads of research papers that underscores the importance of the human connection and "its limitation in replicating genuine human interaction and understanding complex contextual nuances in language use." (Laksana et al., 2024; p. 412). Other research supports human connection when integrating CAI.

Both the ChatGPT and MECAI in MR groups reflected on AI's capability to help with vocabulary development and expansion in areas where the teacher may not be able to help. This aligns with Lie et al. (2023), who underscored the potential of LLMs in providing new vocabulary words in different contextualized situations.

The experimental groups also reflected on CAI use, whether immersive or non-immersive, in terms of their personalized scaffolding capabilities, repeated practice, and immediate personalized and adaptive feedback. This aligns with previous studies that have shown the potential of AI for adaptive, personalized feedback that can enhance students' fluency and accuracy (e.g., 2025; Saraswati et al. 2023; Annamalai et al., 2023; Khailifa & Ginting, 2024; Kholis, 2021; Moussalli & Cardoso, 2020). Qiao and Zhao (2023) found that personalized adaptive feedback significantly enhances students' speaking performance through their ability to adjust task difficulty and provide error correction. Learners' reflections on CAI indicated that it created a safe environment for practicing speaking, which encouraged students to practice with AI without fear of judgement. This is aligned with the research by Ericsson and Johansson (2023) and Ericsson et al. (2023), who showed that AI chatbots and conversational agents helped learners practice language skills without fear of negative evaluation.

Challenges in AI integration

Students noted challenges in AI integration, which aligns with prior research. One of the challenges is the technical limitations of AI. This involves students' low proficiency levels or device technology inconsistency. To explain, low proficiency levels can determine how the students use prompt engineering to be able to converse well with AI

The study also contrasts with previous research (Laksana et al., 2024) that showed overreliance on technology, where most of the students across both AI conditions highlighted integrating the teacher's feedback in terms of responding to specific job interview questions, such as question number 6. Almost all students gave the same response that the teacher gave them during classroom instruction. This emphasizes the importance of human instruction, connection, and interaction in the ESP context which aligns with Nam and Nguyen (2022), Laksana et al. (2024).

The Role of Social Presence and Multimodal Interaction in Mixed Reality

The MECAI in MR group expressed factors such as realism, being there with another, enhancing the feeling of co-presence and immersion (Yuan & Gao, 2023; Wang et al., 2024). Students positively noted that the immersive AI agent in MR has responsive facial animation and gaze tracking combined with the knowledge bank that includes instructions about content generation that the avatar should follow through the LLM created a multimodal immersive communicative experience (Wang et al., 2022; Wang et al., 2024). This aligns with previous studies that showed that such intelligent agents in immersive environments can encourage students to experiment with the language without fear (Wang et al., 2024). The qualitative data aligns with previous research that underscores the design of intelligent agents that have a human-like presence to enhance learner experience (Wang et al., 2024). Moreover, students' perceptions about the teacher integrating and teaching students how to use and interact with CAI aligns with previous studies that show that integrating technology tools in the classroom for effective instruction is vital for the classroom (Mohamed, 2024).

The qualitative findings also contrast previous studies that used rule-based conversational agents, chatbots that followed a fixed set of rules, and exhibited technological limitations such as lack of context understanding, memory, emotional support, and comprehension (Huang et al., 2021).

AI and the Augmented Teacher Role

Students in both AI conditions considered AI as an extension of teacher and curriculum knowledge. Knowing that this context tackles a difficult angle of language learning where the teacher must accommodate both linguistic and professional knowledge, AI in this situation can be a suitable tool to be integrated within a sound pedagogical framework to cater to the students' needs. This aligns with research on intelligence augmentation (Dede et al., 2021)

The Role of Human Connection

All conditions acknowledged the teacher's central pedagogical, emotional, and instructional role in maintaining human connection. A few students in the traditional learning groups preferred teacher instruction over AI for providing attention, feedback, and sensing students' struggles from their facial expressions. However, they acknowledged that they would use AI for more practice in an environment that exhibited less anxiety and was free of judgement. Although the students' responses in the traditional group may suggest bias or social desirability, the researcher ensured to follow Bergen and Labonté's (2020) techniques in limiting social desirability in the reflection interviews by providing clear, explainable, direct, and contextualized questions. Traditional students also seemed to have a competitive edge when preparing for their posttests and relied on memorizing responses for the posttest. This reliance on rote learning was likely because they lacked personalized practice like the other experimental groups, in which they could negotiate meaning and use critical thinking skills (Abbas & Harrison, 2025). On the other hand, students in both AI conditions acknowledged the necessity of a human instructor's role in providing constructive feedback, being patient, emotional support, and culturally informed explanations while integrating AI to supplement teacher and curriculum knowledge. These are factors that research in language learning describes as irreplaceable in the language learning classroom (Rodrigues, 2025; Tolba et al., 2024).

Limitations

This study has several limitations. This study examined only one level of proficiency, namely, the high beginner level. It would have been of utmost importance to explore how different levels interact with AI and how their anxiety changes over time. Studies also recommend exploring and comparing different proficiency levels to understand the differences in language acquisition and development (Zhang, 2019).

Despite students' interest in acquiring HMDs and using them for learning, these devices are still expensive for community college students. However, ChatGPT has other advantages in terms of ubiquity and accessibility.

Implications

This study aims to inform researchers, educators, and language-learning curriculum and instructional designers about the importance of supplementing high-stakes language-learning activities, such as job interviews, with multimodal embodied conversational AI. The advent of LLMs has provided educators with a wealth of resources that can complement teachers knowledge and curriculum limitations. The most important factor is AI literacy and how students can understand prompt engineering as a tool for facilitating learning. Regarding MR HMDs, initially, the HCCC coordinator and the instructor who allowed me to collect the data in his classroom were skeptical, but when they experienced how MECAI could provide guidance and examples, they became interested in the study.

For ESL educators teaching ESP, multimodal generative AI tools can be used for conversation analytics and feedback to students. Conversation analytics turn talk into data. Teachers can examine students' interactions and analyze negotiation patterns and quality of AI feedback. Teachers can also use AI tools to bridge the gap between traditional classroom learning and real-world scenarios, by using role-play activities that can engage students and create a judgement free environment. AI can also supplement curriculum by identifying discipline-specific language gaps and generating authentic discourse scenarios. This in turn would underscore the need for ESL programs to provide professional development training for topics related to prompt engineering, integrating AI tools within sound pedagogical frameworks, and providing grants to encourage educators to explore ways to integrate AI into instruction. Instructional designers can embed human-in-the -loop mechanisms that require human review, tailoring integration of AI within sound teaching methods, and iteratively refine role-play activities. This can have a good impact on students' learning when it comes to designing AI role-play simulations, designing activities that focus on form and content, and monitoring AI feedback and students' interactions for optimal learning.

The implications of this research extend beyond ESL and ESP to other industries and subject matters. Embodied conversational AI affordances include aligning AI output with curriculum goals, creating multimodal conversational AI agents, and immersive environments that can mimic real-life situations, such as interfaith dialogues, community problem-solving, and conflict-sensitive discussions, where AI enables learners to exchange ideas and engage in meaningful dialogues. These technologies can create psychologically safe, low-pressure environments in which participants explore perspectives, reflect, and communicate without fear of judgement.

As many institutions are leading curriculum innovation and integrating artificial intelligence in the classroom, this research offers some insights through its mixed method approach to understand how AI can support students' learning and reduce their anxiety.

Future Work

There are several areas in this study that I would like to revisit and revise. First, the sample size should be increased and the unequal allocation of groups should be addressed. I would extend the duration of the intervention to understand whether an extended duration of practice would decrease students' anxiety. I would also include another data collection method for the reflection interviews, such as Linguistic Inquiry and Word Count (LIWC). This computational text analysis tool can help in understanding respondents' language pattern use in terms of sentiment analysis, cognitive processes, pronouns, social orientation, and other aspects that can provide a rich interpretation of students' perceptions and challenges during the interaction. I would measure students' heart rates while they interact with AI and use heart rate during pre- and posttests. Conducting a longitudinal study with the same group of students across different levels can help track students' speaking development and anxiety reduction over time.

Conclusion

Multimodal Embodied Artificial Intelligence is a technology that has been difficult to employ in the past because of its cost and hardware requirements. With the advent of large language models and HMDs becoming less expensive, this breakthrough in technology will provide opportunities for language learning researchers to explore the impact of artificial intelligence on students' learning and their affect. This study found that integrating MECAI into MR can have a significant effect on students' performance. The ChatGPT voice application also had a good impact in comparison with the control group that memorized responses. Although no significant difference in anxiety was found, students reflection interview responses underscored two main themes: AI is crucial to providing a judgement-free environment. Second, human teachers are necessary in the classroom for human connection that can help reduce anxiety.

References

Abbas, Q., & Harrison, R. (2025). ChatGPT and generative AI: Addressing critical thinking in education's transition from "what" to "why".

Abbasi, A. M., Khalil, A., & John, S. (2019). ESL learners' performance correlates of social and academic anxiety. *ELF Annual Research Journal, 21*(1), 14–81.

Abdiganieva, N., & Shamuratov, Z. (2022). The importance of spoken interaction in EFL classrooms. *Renaissance in the Paradigm of Innovations in Education and Technology in the 21st Century, 1*, 216–217.

Abowitz, D. A., & Toole, T. M. (2010). Mixed method research: Fundamental issues of design, validity, and reliability in construction research. *Journal of Construction Engineering and Management, 136*(1), 108–116. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000026

Achivida, J. J. (2025). Multimodal language teaching and language skills development with AI and VR integration: A systematic literature review. *Psychology and Education: A Multidisciplinary Journal, 31*(6), 587–597. https://doi.org/10.5281/zenodo.14797458

Ahmad Bukhari, N. I., Ismail, L., Abu Kassim, N. L., Razali, A. B., Noordin, N., & Mohd Noh, M. F. (2023). Topic familiarity effects on performance in speaking assessment tasks. *Arab World English Journal, 14*(4), 213–232. https://doi.org/10.24093/awej/vol14no4.13

Ahmed, M. K. (2018). Pedagogy in speaking: Challenges addressed by teacher-student in the ESL context. *International Journal of Applied Linguistics and English Literature, 7*(3), 97–107. https://doi.org/10.7575/aiac.ijalel.v.7n.3p.97

Ahn, S. K. (2018). *Korean EFL college students' acceptance and use of smartphone applications for English language learning* [Doctoral dissertation, Alliant International University]. ProQuest Dissertations and Theses Global.

Al-Khasawneh, F. M. (2023). From text to tech: Investigating chatbots' role in enhancing oral proficiency in second language learners. *Journal of Southwest Jiaotong University, 58*(5), 601–610. https://doi.org/10.35741/issn.0258-2724.58.5.45

AlKhayat, A. (2017). Exploring the effectiveness of using chatbots in the EFL classroom. In *Teaching English Reflectively with Technology* (pp. 20–36).

Allcoat, D., Hatchard, T., Azmat, F., Stansfield, K., Watson, D., & von Mühlenen, A. (2021). Education in the digital age: Learning experience in virtual and mixed realities. *Journal of Educational Computing Research, 59*(5), 795–816. https://doi.org/10.1177/0735633120985120

Altunkaya, H. (2018). Speech anxiety in guidance and psychological counselling department students. *Cypriot Journal of Educational Sciences, 13*(3), 328–339. https://doi.org/10.18844/cjes.v13i3.3764

AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon, 10*(10), e31053. https://doi.org/10.1016/j.heliyon.2024.e31053

Alyahya, S. A. (2020). *Social cues in animated pedagogical agents for second language learners: The application of the embodiment principle in video design* [Doctoral dissertation, University of South Florida]. Digital Commons @ University of South Florida. https://digitalcommons.usf.edu/etd/8724

Amengual-Pizarro, M. (2018). Foreign language classroom anxiety among English for Specific Purposes (ESP) students. *International Journal of English Studies, 18*(2), 145–159. https://doi.org/10.6018/ijes.331111

Ampatuan, R. A., & San Jose, A. E. (2016). Role play as an approach in developing students communicative competence. *International Journal for Innovation Education and Research, 4*(1), 18–24.

Anabuki, M., Kakuta, H., Yamamoto, H., & Tamura, H. (2000, April). Welbo: An embodied conversational agent living in mixed reality space. In *CHI'00 extended abstracts on Human factors in computing systems* (pp. 10–11).

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial intelligence, 149*(1), 91–130.

Ansari, M. S. (2015). Speaking anxiety in ESL/EFL classrooms: A holistic approach and practical study. *International Journal of Educational Investigations, 2*(4), 38–46.

Argyle, M. (2013). *The social psychology of everyday life*. Routledge.

Arham, R., Yassi, A. H., & Arafah, B. (2016). The use of role play to improve teaching speaking. *International Journal of Scientific and Research Publications, 6*(3), 239–241.

Aydın, B. (1999). *A study of sources of foreign language classroom anxiety in speaking and writing classes*. Anadolu University (Turkey).

Ayres, J., & Crosby, S. (1995). Two studies concerning the predictive validity of the personal report of communication apprehension in employment interviews. *Communication Research Reports, 12*(2), 145–151.

Ayuningtyas, P., Mauludin, L. A., & Prasetyo, G. (2022). Investigating the anxiety factors among English for Specific Purposes students in a vocational education setting. *Language Related Research, 13*(3), 31–54.

Aziz, A. A., & Kashinathan, S. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *Development, 10*(2), 983–991.

Bacon, S. M., & Finnemann, M. D. (1990). A study of the attitudes, motives, and strategies of university foreign language students and their disposition to authentic oral and written input. *The Modern Language Journal, 74*(4), 459–473.

Bączkowska, A. (2021, December). An overview of popular website platforms and mobile apps for language learning. *Forum Filologiczne Ateneum, 9*(1), 9–35. Akademia Ateneum w Gdańsku.

Bao, M. (2019). Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety: Investigation of a concept. *Arab World English Journal, 5*(5), 28–40.

Baralt, M., & Gurzynski-Weiss, L. (2011). Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Language Teaching Research, 15*(2), 201–229.

Beatty, K. (2013). *Teaching & researching: Computer-assisted language learning*. Routledge.

Beauvois, M. H. (1992). Computer-assisted classroom discussion in the foreign language classroom: Conversation in slow motion. *Foreign Language Annals, 25*(5), 455–464.

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427.

Benati, A. G. (2009). *Japanese Language Teaching*. Continuum.

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.

Bibauw, S., Van den Noortgate, W., François, T., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology, 26*(1), 1–24.

Billinghurst, M., Kato, H., & Poupyrev, I. (2002). The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications, 21*(3), 6–8.

Bin-Tahir, S. Z., & Hanapi, H. (2017). Lecturers' Method in Teaching Speaking at the University of Iqra Buru.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 120–136.

Borup, J., West, R. E., Thomas, R., & Graham, C. R. (2014). Examining the impact of video feedback on instructor social presence in blended courses. *The International Review of Research in Open and Distributed Learning, 15*(3).

Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Brown, S. A., Fuller, R. M., & Vician, C. (2004). Who's afraid of the virtual world? Anxiety and computer-mediated communication. *Journal of the Association for Information Systems, 5*(2), 2.

Burns, A. (2012). A holistic approach to teaching speaking in the language classroom. In *Symposium* (pp. 165–178).

Cabero-Almenara, J., Barroso-Osuna, J., & Martinez-Roig, R. (2021). Mixed, augmented and virtual, reality applied to the teaching of mathematics for architects. *Applied Sciences, 11*(15), 7125.

Cai, Q., Lin, Y., & Yu, Z. (2024). Factors influencing learner attitudes towards ChatGPT-assisted language learning in higher education. *International Journal of Human–Computer Interaction, 40*(22), 7112–7126.

Campos, A. A. M. (2017). *Adopting smartphone applications for Second Language Acquisition: investigating readiness and acceptance of mobile learning in two Higher Education Institutions* (Doctoral dissertation, Universidade NOVA de Lisboa).

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.

Cantos, K. F. S., Pintado, J. A. C., Coello, M. B. B., & Castillo, L. A. C. (2024). Strategies for Reducing Anxiety in EFL Classroom: An Academic Review. *Ciencia Latina Revista Científica Multidisciplinar, 8*(2), 7159–7169.

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., ... & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders, 61*, 27–36.

Carless, D. (2008). Student use of the mother tongue in the task-based classroom. *ELT Journal, 62*(4), 331–338.

Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System, 31*(4), 485–500.

Carver, C. S. (1979). A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology, 37*(8), 1251.

Carver, C. S., & Scheier, M. F. (2012). *Attention and self-regulation: A control-theory approach to human behavior*. Springer Science & Business Media.

Cassell, J. (2001). Embodied conversational agents: representation and intelligence in user interfaces. *AI Magazine, 22*(4), 67–67.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., & Yan, H. (2000). Designing embodied conversational agents. *Embodied Conversational Agents, 29*.

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. *Intercultural Language Use and Language Learning*, 41–57.

Chapelle, C. A. (2006). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. John Benjamins Publishing Company.

Chen, H. (2018). The use of 3D holographic platform for L2 learning. *Journal of Educational Technology & Society, 21*(3), 1–13.

Chen, H. I., & Sevilla-Pavón, A. (2023). Negotiation of meaning via virtual exchange in immersive virtual reality environments.

Chen, S. S. C., & Duh, H. (2019). Interface of mixed reality: From the past to the future. *CCF Transactions on Pervasive Computing and Interaction, 1*(1), 69–87.

Chen, Y. (2018, August). Reducing language speaking anxiety among adult EFL learners with interactive holographic learning support system. In *International conference on innovative technologies and learning* (pp. 101–110). Cham: Springer International Publishing.

Chesham, R. K., Malouff, J. M., & Schutte, N. S. (2018). Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety. *Behaviour Change, 35*(3), 152–166.

Chiu, T. K., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments, 32*(7), 3240–3256.

Chomsky, N. (2014). *Aspects of the Theory of Syntax*. MIT press.

Christoforou, M., Thrasher, T., & Ijiri, A. (2024, December). Technology-mediated task-based language teaching in a virtual reality learning environment. In *Polytechnic University of Valencia Congress, EuroCALL 2024: CALL for humanity*. https://doi.org/10.4995/EuroCALL2024.2024.19028

Christoforou, M., Thrasher, T., & Ijiri, A. (2024). Virtual reality-mediated language learning: A case study of immersive learning in the metaverse. In R. Vurdien & W. Chambers (Eds.), *Technology-mediated language learning and teaching*. IGI Global. https://doi.org/10.4018/979-8-3693-2687-9

Christoforou, M., Xerou, E., & Papadima-Sophocleous, S. (2019). Integrating a virtual reality application to simulate situated learning experiences in a foreign language course. *CALL and Complexity–Short Papers from EUROCALL*, 82–87.

Chuah, J. H., Lok, B., & Black, E. (2013). Applying mixed reality to simulate vulnerable populations for practicing clinical communication skills. *IEEE Transactions on Visualization and Computer Graphics, 19*(4), 539–546.

Churchill, D., Chiu, T. K., & Gu, N. J. (2014). Mobile Learning, MOOCs and 21st Century Learning.

Clark, D. M., & Wells, A. (1995). A cognitive model of social phobia. In R. G. Heimberg, M. R. Liebowitz, D. A. Hope, & F. R. Schneier (Eds.), *Social phobia: Diagnosis, assessment and treatment* (pp. 69–93). Guilford Press.

Cohen, A. D. (2020). Considerations in assessing pragmatic appropriateness in spoken language. *Language Teaching, 53*(2), 183–202.

Coniam, D. (2014). The linguistic accuracy of chatbots: usability from an ESL perspective. *Text & Talk, 34*(5), 545–567.

Cook, K. W., Vance, C. A., & Spector, P. E. (2000). The relation of candidate personality with selection-interview outcomes. *Journal of Applied Social Psychology, 30*(4), 867–885.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571.

Dang, B., Huynh, L., Gul, F., Rosé, C., Järvelä, S., & Nguyen, A. (2025). Human–AI collaborative learning in mixed reality: Examining the cognitive and socio-emotional interactions. *British Journal of Educational Technology*.

Dede, C., Etemadi, A., & Forshaw, T. (2021). *Intelligence augmentation: Upskilling humans to complement AI*. The Next Level Lab at the Harvard Graduate School of Education. President and Fellows of Harvard College: Cambridge, MA.

Denizci Nazligul, M., Yilmaz, M., Gulec, U., Yilmaz, A. E., Isler, V., O'Connor, R. V., ... & Clarke, P. (2019). Interactive three-dimensional virtual environment to reduce the public speaking anxiety levels of novice software engineers. *IET Software, 13*(2), 152–158.

Dentella, V., Günther, F., Murphy, E., Marcus, G., & Leivada, E. (2024). Testing AI on language comprehension tasks reveals insensitivity to underlying meaning. *Scientific Reports, 14*(1), 28083.

Dewaele, J. M., & Dewaele, L. (2017). The dynamic interactions in foreign language classroom anxiety and foreign language enjoyment of pupils aged 12 to 18. A pseudo-longitudinal investigation. *Journal of the European Second Language Association, 1*(1), 12–22.

Dieterle, E., Dede, C., & Schrier, K. (2007). "Neomillennial" Learning Styles Propagated by Wireless Handheld Devices. In *Ubiquitous and pervasive knowledge and learning management: Semantics, social networking and new media to their full potential* (pp. 35–66). IGI Global Scientific Publishing.

Dikmen, M. (2021). EFL learners' foreign language learning anxiety and language performance: A meta-analysis study. *International Journal of Contemporary Educational Research, 8*(3), 206–222.

Dincer, B., Özçelik, S. K., Özer, Z., & Bahçecik, N. (2022). Breathing therapy and emotional freedom techniques on public speaking anxiety in Turkish nursing students: A randomized controlled study. *Explore, 18*(2), 226–233.

Ding, D., & Yusof, M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study. *Humanities and Social Sciences Communications, 12*(1), 1–16.

Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *Tesol Journal, 8*(4), 811–830.

Dodigovic, M. (2007). Artificial intelligence and second language learning: An efficient approach to error remediation. *Language Awareness, 16*(2), 99–113.

Doswell, J. T. (2005). It's virtually pedagogical: pedagogical agents in mixed reality learning environments. In *ACM SIGGRAPH 2005 educators program* (pp. 25-es).

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes*. Cambridge University Press.

Ebadi, S., & Amini, A. (2024). 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, 32*(2), 655–673.

Edelmann, R. J., & Baker, S. R. (2002). Self-reported and actual physiological responses in social phobia. *British Journal of Clinical Psychology, 41*(1), 1–14.

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), e12667.

Elahemer, H. H. M., & Said, S. B. H. M. (2022). A Systematic Review of Interventions Used to Reduce University Students' Speaking Anxiety. *International Journal of Education and Practice, 10*(2), 128–149.

Ellis, R. (2003). Task-based Language Learning and Teaching. *Asian Journal of English Language Teaching, 13*, 125–129.

Ellis, R. (2004). *Task-based language teaching and learning*. Oxford University Press.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*(2), 336.

Farah, J. C., Sharma, V., Ingram, S., & Gillet, D. (2021, November). Conveying the perception of humor arising from ambiguous grammatical constructs in human-chatbot interaction. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (pp. 257–262).

Fathi, J., & Mohammaddokht, F. (2021). Foreign language enjoyment and anxiety as the correlates of the ideal L2 self in the English as a foreign language context. *Frontiers in Psychology, 12*, 790648.

Feiler, A. R., & Powell, D. M. (2013). Interview anxiety across the sexes: Support for the sex-linked anxiety coping theory. *Personality and Individual Differences, 54*(1), 12–17.

Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *Calico Journal, 19*(2), 279–284.

Ferris, D., & Tagg, T. (1996). Academic listening/speaking tasks for ESL students: Problems, suggestions, and implications. *TESOL Quarterly, 30*(2), 297–320.

Fitriyanti, I. D., & Malihah, N. (2024). The Problems Encountered by English Foreign Language Learners in English Speaking Skill: Systematic Literature Review. *Education And Linguistics Knowledge Journal, 6*(2), 170–181.

Foster, P., & Ohta, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics, 26*(3), 402–430.

Fremouw, W. J., & Zitter, R. E. (1978). A comparison of skills training and cognitive restructuring-relaxation for the treatment of speech anxiety. *Behavior Therapy, 9*(2), 248–259.

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior, 93*, 279–289.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools.

Fryer, L., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions.

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018, October). Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. In *International conference on internet science* (pp. 145–156). Cham: Springer International Publishing.

Fuente, M. J. (2003). Is SLA interactionist theory relevant to CALL? A study on the effects of computer-mediated interaction in L2 vocabulary acquisition. *Computer Assisted Language Learning, 16*(1), 47–81.

Gamper, J., & Knapp, J. (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning, 15*(4), 329–342.

Gan, Z. (2013). Understanding English speaking difficulties: An investigation of two Chinese populations. *Journal of Multilingual and Multicultural Development, 34*(3), 231–248.

Gandeel, A. M. (2016). *English language teachers' beliefs and practices regarding the teaching of speaking* (Doctoral dissertation, University of Leeds).

Ganster, D. C., Crain, T. L., & Brossoit, R. M. (2018). Physiological measurement in the organizational sciences: A review and recommendations for future use. *Annual Review of Organizational Psychology and Organizational Behavior, 5*(1), 267–293.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology, 119*(7), 1525–1532.

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal, 58*(4), 363–374.

Gladstone, J. R., Schroeder, N. L., Heidig, S., Zhang, S., Palaguachi, C., & Pitera, M. (2025). Do Pedagogical Agents Enhance Student Motivation? Unraveling the Evidence Through

Meta-Analysis. *Educational Psychology Review, 37*(3), 72. https://doi.org/10.1007/s10648-025-10050-2

Godwin-Jones, R. (2021). Evolving technologies for language learning.

Goktas, Y. (2023). *Mobile learning of English speaking and listening skills via smartphones* (Doctoral dissertation, University of Birmingham).

Gonzales, D. M., Kalamkar, S., Jörg, S., & Grubert, J. (2025). Behavioral and Symbolic Fillers as Delay Mitigation for Embodied Conversational Agents in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*.

González-Lloret, M. (2003). Designing task-based CALL to promote interaction: En busca de esmeraldas.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking, 14*(3), 99–105.

Gorini, A., Griez, E., Petrova, A., & Riva, G. (2010). Assessment of the emotional responses produced by exposure to real food, virtual food and photographs of food in patients affected by eating disorders. *Annals of General Psychiatry, 9*(1), 30.

Gosiewska-Turek, B. (2018). The impact of anxiety on speaking in adolescent and adult groups of English learners. *The Journal of Education, Culture, and Society, 9*(2), 93–106.

Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal, 98*(2), 574–588.

Gripse, J., & Sellerberg, V. (2025). *What is the role of digital tools in combination with task-based language practices for language learners, and how does it affect learning* *outcomes?* [Vilken roll har digitala verktyg i kombination med uppgiftsbaserad språkpraxis för språkinlärare, och hur påverkar det läranderesultat?].

Gumperz, J. J. (1992). Interviewing in intercultural situations. *Talk at Work*.

Gumperz, J. J. (Ed.). (1982). *Language and social identity* (No. 2). Cambridge University Press.

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education, 11*(3), 8–26.

Hajiyeva, B. (2024). Language Anxiety in ESL Learners: Causes, Effects, and Mitigation Strategies. *EuroGlobal Journal of Linguistics and Language Education, 1*(1), 119–133.

Hammad, E. A., & Ghali, E. M. A. (2015). Speaking anxiety level of Gaza EFL pre-service teachers: Reasons and sources. *World Journal of English Language, 5*(3), 52.

Han, D., & Ryu, J. (2024, March). Effects of AI Chatbot-based Activities on Korean EFL Learners' Speaking Ability and Affective Domain. In *Society for Information Technology & Teacher Education International Conference* (pp. 1616–1620). Association for the Advancement of Computing in Education (AACE).

Han, D., & Ryu, J. (2025). Exploring Embodied Conversational Agents in Extended Reality for Language Learning. *Immersive Learning Research-Practitioner*, 38–42.

Han, E., & Bailenson, J. N. (2024). Social interaction in VR. In *Oxford Research Encyclopedia of Communication*.

Han, T., & Keskin, F. (2016). Using a mobile application (WhatsApp) to reduce EFL speaking anxiety. *Gist: Education and Learning Research Journal*, (12), 29–50.

Handayani, F. (2016, May). Instagram as a teaching tool? Really. In *Proceedings of the fourth international seminar on English language and teaching (ISELT-4)* (Vol. 4, No. 1, pp. 320–327).

Hapsari, I. P., & Wu, T. T. (2022, August). AI chatbots learning model in English speaking skill: Alleviating speaking anxiety, boosting enjoyment, and fostering critical thinking. In *International conference on innovative technologies and learning* (pp. 444–453). Cham: Springer International Publishing.

Hardy, L. (1999). Stress, anxiety and performance. *Journal of Science and Medicine in Sport, 2*(3), 227–233.

Harmer, J. (2001). *The practice of English language teaching*. London/New York.

Hartholt, A., Mozgai, S., & Rizzo, A. S. (2019, July). Virtual job interviewing practice for high-anxiety populations. In *Proceedings of the 19th ACM international conference on intelligent virtual agents* (pp. 238–240).

Hassanein, K., & Head, M. (2007). Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies, 65*(8), 689–708.

Heerey, E. A., & Kring, A. M. (2007). Interpersonal consequences of social anxiety. *Journal of Abnormal Psychology, 116*(1), 125.

Heift, T., & Schulze, M. (2015). Tutorial computer-assisted language learning. *Language Teaching, 48*(4), 471–490.

Hein, R. M., Wienrich, C., & Latoschik, M. E. (2021). A systematic review of foreign language learning with immersive technologies (2001-2020). *AIMS Electronics and Electrical Engineering, 5*(2), 117–145.

Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005, June). Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ innovate learning* (pp. 3913–3918). Association for the Advancement of Computing in Education (AACE).

Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments, 5*(3), 290–301.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *Tesol Quarterly, 40*(1), 109–131.

Hismanoglu, M. (2012). Integrating second life into an EFL classroom: A new dimension in foreign language learning and teaching. *International Journal on New Trends in Education & their Implications (IJONTE), 3*(4).

Hoang, N. T., Han, D. N., & Le, D. H. (2023). Exploring Chatbot AI in improving vocational students' English pronunciation. *AsiaCALL Online Journal, 14*(2), 140–155.

Höl, D., & Kasımi, Y. (2022). A growing dilemma: English speaking anxiety-in-EFL classrooms: A review of research. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (28), 421–438.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132.

Hosseini, S., Quan, J., Deng, X., Miyake, Y., & Nozawa, T. (2024). Avatar-Based Feedback in Job Interview Training Impacts Action Identities and Anxiety. *IEEE Transactions on Affective Computing, 15*(3), 1608–1620.

Huang, C. Y., Yang, M. C., Huang, C. Y., Chen, Y. J., Wu, M. L., & Chen, K. W. (2018, December). A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. In *2018 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 1791–1795). IEEE.

Huang, H., Chen, Y., & Rau, P. L. P. (2022). Exploring acceptance of intelligent tutoring system with pedagogical agent among high school students. *Universal Access in the Information Society, 21*(2), 381–392.

Huang, J. (2010). Grammar instruction for adult English language learners: A task-based learning framework. *Journal of Adult Education, 39*(1), 29–37.

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.

Hubbard, P. (2021). An invitation to CALL: Foundations of computer-assisted language learning. *Asia-Pacific Association for Computer-Assisted Language Learning (APACALL)*.

Hubbard, P., & Levy, M. (2016). The scope of CALL education: Language teacher education and development. In F. Farr & L. Murray (Eds.), *The Routledge Handbook of Language Learning and Technology* (pp. 303–317).

Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. *Human Resource Management Review, 21*, 353–367.

Hughes, R., & Reed, B. S. (2016). *Teaching and researching speaking* (2nd ed.). Routledge.

Hunter, D. (2009). *Communicative Language Teaching and the ELT Journal: A corpus-based approach to the history of a discourse* [Doctoral dissertation, University of Warwick]. Warwick Research Archive Portal. https://wrap.warwick.ac.uk/3731/

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin.

Ironsi, C. S. (2023). Investigating the use of virtual reality to improve speaking skills: Insights from students and teachers. *Smart Learning Environments, 10*(1), 53.

Iwasaki, N., & Oliver, R. (2003). Chat-line interaction and negative feedback. *Australian Review of Applied Linguistics, 17*, 60–73.

Jensen, C. J., & Cadierno, T. (2022). Differences in mobile-assisted acquisition of receptive and productive vocabulary knowledge: A case study using Mondly. *The Language Learning Journal*, 1–16.

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*, Article 104898.

Jepson, K. (2005). Conversations and negotiated interactions in text and voice chat rooms. *Language Learning & Technology, 9*, 79–98.

Jerald, J. (2016). *The VR book: Human-centered design for virtual reality*. Association for Computing Machinery and Morgan & Claypool.

Jiangqin, X., Feng, L., & Min, J. (2008). Task-based language teaching. *International Conference on Computer Science & Software Engineering*.

Jung, H., Kim, H. J., So, S., Kim, J., & Oh, C. (2019, May). TurtleTalk: An educational programming game for children with voice user interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6).

Kana, F. (2015). Investigation of preservice teachers' speech anxiety with different points of view. *International Journal of Progressive Education, 11*(3), 140–152.

Kang, H. (2022). Effects of Artificial Intelligence (AI) and native speaker interlocutors on ESL learners' speaking ability and affective aspects. *Multimedia-Assisted Language Learning, 25*(2).

Kaplan-Rakowski, R., & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety. *Smart Learning Environments, 10*, Article 46. https://doi.org/10.1186/s40561-023-00263-9

Kaur, M. (2022). Investigation of factors influencing Malaysian ESL learners' English speaking skills. *International Journal of Advanced Research in Education and Society, 4*(2), 1–6.

Keck, C., Iberri-Shea, G., Tracy-Ventura, N., & Wa-Mbaleka, S. (2006). Task-based interaction and CALL. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 91–131). John Benjamins.

Kerekes, J. A. (2006). Winning an interviewer's trust in a gatekeeping encounter. *Language in Society, 35*(1), 27–57.

Khamkhien, A. (2010). Thai learners' English pronunciation competence: Lesson learned from word stress assignment. *Journal of Language Teaching and Research, 1*(6), 757–764. https://doi.org/10.4304/jltr.1.6.757-764

Khany, R., & Kamalvand, A. (2022a). 100 years of research on English language learning/teaching materials: A systematic literature review. *TESL-EJ, 26*(4).

Khany, R., & Kamalvand, A. (2022b). One hundred years of ELT materials: A review. *TESL-EJ, 25*(4).

Kikalishvili, S. (2023). Unlocking the potential of GPT-3 in education: Opportunities, limitations, and recommendations for effective integration. *Interactive Learning Environments*, 1–13.

Kilic, N., Eryilmaz, A., & Dinç, S. Y. (2018). Effectiveness of psychoeducational group training on Turkish EFL learners' English speaking anxiety. *Journal of Hacettepe University Faculty of Education, 33*(4), 1020–1037.

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments, 21*(4), 373–387.

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

Kim, N. Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning, 19*(4).

Kim, N. Y. (2018). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence, 16*(8).

Kim, S., & Lee, J. M. (2023). The effect of voice-enabled ChatGPT on negotiation of meaning of Korean EFL learners. *Applied Linguistics, 39*(3), 3–33.

Kim, Y., & Baylor, A. L. (2007). Pedagogical agents as social models to influence learner attitudes. *Educational Technology*, 23–28.

Kim, Y.-H., Hwang, K.-H., & Cho, O.-H. (2018). Simulation education with problem-based learning: Effect on nursing students' communication apprehension. *Social Behavior and Personality: An International Journal, 46*(1), 151–160.

Kiruthiga, E., & Christopher, G. (2022). The impact of affective factors in English speaking skills. *Theory and Practice in Language Studies, 12*(12), 2478–2485.

Kita, T., Nagaoka, C., Hiraoka, N., Suzuki, K., & Dougiamas, M. (2018, March). A discussion on effective implementation and prototyping of voice user interfaces for learning activities on Moodle. In *CSEDU (1)* (pp. 398–404).

Koç, F. Ş. (2024). *The development of listening and speaking skills in EFL via an artificially intelligent chatbot application: A quasi-experimental design study* [Doctoral dissertation, Middle East Technical University].

Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation, 17*(1–2), 214–226.

Kötter, M. (2003). Negotiation of meaning and code-switching in online tandems. *Language Learning & Technology, 7*, 145–172.

Krämer, N. C., Kopp, S., Jordan, P., & Wrede, S. (2013). The effects of a virtual agent's nonverbal behavior on the perception of its social capabilities and the user's social experience. *Journal of Computer-Mediated Communication, 18*(4), 524–541.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. D. (1983). *The natural approach: Language acquisition in the classroom*. Alemany.

Krebt, D. M. (2017). The effectiveness of role play techniques in teaching speaking for EFL college students. *Journal of Language Teaching and Research, 8*(5), 863.

Kühne, F., & Wge, M. (2019). The effect of a virtual agent's appearance on foreign language speaking practice. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (pp. 165–167).

Kuppusamy, M., Kamaldeen, D., Pitani, R., Amaldas, J., & Shanmugam, P. (2018). Effects of Bhramari Pranayama on health–a systematic review. *Journal of Traditional and Complementary Medicine, 8*(1), 11–16.

Kyrlitsias, C., & Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality, 2*, 786665.

Laksana, I. P. Y., Virginiya, P. T., & Yuliantini, N. N. (2024, December). Students' perceptions on using AI: Factors, benefits, and challenges in influencing their speaking performances. In *Proceedings of the International Conference on Sustainable Green Tourism Applied Science-Social Applied Science 2024 (ICoSTAS-SAS 2024)* (Vol. 308, p. 407). Springer Nature.

Lambert, C. (2010). *A task-based needs analysis: Putting principles into practice* [Unpublished research paper]. University of Kitakyushu.

Lan, Y.-J. (2020). Immersion, interaction, and collaboration in virtual environments for language learning. *Language Learning & Technology, 24*(1), 36–54.

Latkin, C. A., Edwards, C., Davey-Rothwell, M. A., & Tobin, K. E. (2017). The relationship between social desirability bias and self-reported health behaviors in a community-based population. *PLOS ONE, 12*(9), e0183769.

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *State-Trait Anxiety Inventory (German version)*. Beltz.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin, 32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lee, D., & Kim, S.-Y. (2018). Effects of input- and output-based planning on Korean college students' communication strategy and speaking anxiety. *Journal of Asia TEFL, 15*(4), 959–975.

Lee, H., Hsia, C. C., Tsoy, A., Choi, S., Hou, H., & Ni, S. (2023). VisionARy: Exploratory research on contextual language learning using AR glasses with ChatGPT. In *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter* (pp. 1–6).

Lee, J. S. (2020). Chatbots as language learning partners: Effects on speaking performance and interactional competence. *Computer Assisted Language Learning, 33*(3), 271–294.

Lee, J., Lee, D., & Lee, J. G. (2022). Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. *International Journal of Human–Computer Interaction*, 1–12.

Lee, K. (2002). *Task-based learning for pronunciation* [Unpublished research paper]. University of Canberra.

Lee, L. (2001). Online interaction: Negotiation of meaning and strategies used among learners of Spanish. *ReCALL, 13*, 232–244.

Lee, L. (2007). Fostering second language oral communication through constructivist interaction in desktop videoconferencing. *Foreign Language Annals, 40*, 635–649.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2009). Avatar e-mail versus traditional e-mail: Perceptual difference and media selection difference. *Decision Support Systems, 46*(2), 451–467.

LeFebvre, L., LeFebvre, L. E., Allen, M., Buckner, M. M., & Griffin, D. (2020). Metamorphosis of public speaking anxiety: Student fear transformation throughout the introductory communication course. *Communication Studies, 71*(1), 98–111.

León, W. U., & Cely, E. V. (2010). Encouraging teenagers to improve speaking skills through games in a Colombian public school. *PROFILE: Issues in Teachers' Professional Development, 12*(1), 11–31.

Li, B., Bonk, C. J., & Kou, X. (2023). Exploring the multilingual applications of ChatGPT: Uncovering language learning affordances in YouTuber videos. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 13*(1), 1–22. https://doi.org/10.4018/IJCALLT.326135

Li, H., Zhang, J., Ding, L., & Wang, G. (2025). The facts about the effects of pedagogical agents on learners' cognitive load: a meta-analysis based on 24 studies. *Frontiers in Psychology, 16*, 1635465. https://doi.org/10.3389/fpsyg.2025.1635465

Liaw, S. Y., Tan, J. Z., Lim, S., Zhou, W., Yap, J., Ratan, R., Ooi, S. L., Wong, S. J., Seah, B., & Chua, W. L. (2023). Artificial intelligence in virtual reality simulation for interprofessional communication training: Mixed method study. *Nurse Education Today, 122*, 105718.

Liew, T. W., & Tan, S. M. (2016). The effects of pedagogical agent's persona and realism in a multimedia learning environment. *Computers & Education, 94*, 61–76.

Lim, S., Schmälzle, R., & Bente, G. (2025). Artificial social influence via human-embodied AI agent interaction in immersive virtual reality (VR): Effects of similarity-matching during health conversations. *Computers in Human Behavior: Artificial Humans*, Article 100172.

Lin, S.-W., & Bollen, D. H. (2024). From anxiety to engagement: Transforming an English for Specific Purposes classroom through digital storytelling. *Studies in English Language and Education*.

Linh, P. M., Starčič, A. I., & Wu, T. T. (2022, August). Challenges and opportunities of education in the COVID-19 pandemic: Teacher perception on applying AI chatbot for online language learning. In *International Conference on Innovative Technologies and Learning* (pp. 501–513). Springer.

Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge University Press.

Littlewood, W. (2006). *Communicative and task-based language teaching in East Asian classrooms* [Unpublished research paper]. Hong Kong Institute of Education.

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*(1), 50.

Lomicka, L., & Lord, G. (2007). Social presence in virtual communities of foreign language (FL) teachers. *System, 35*(2), 208–228.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Lopez-Gazpio, I. (2025). Integrating large language models into accessible and inclusive education: Access democratization and individualized learning enhancement supported by generative artificial intelligence. *Information, 16*(6), 473. https://doi.org/10.3390/info16060473

Lu, L. (2025). *Do L2 speakers' assumptions about the test examiner influence their speaking anxiety in an oral exam? A reverse linguistic stereotyping study* [Doctoral dissertation, Concordia University].

Luo, H. (2013). Foreign language anxiety: Past and future. *Chinese Journal of Applied Linguistics, 36*(4), 442–464.

Mabrito, M. (2024). Artificial intelligence in the classroom: Conversation design and prompt engineering for English majors. *International Journal of Technologies in Learning, 31*(2).

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review, 19*, 203–218.

MacIntyre, P. D., & Gardner, R. C. (1991a). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning, 41*(1), 85–117.

MacIntyre, P. D., & Gardner, R. C. (1991b). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning, 41*(4), 513–534.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408–452). Oxford University Press.

Mahdi, H. S. (2014). The impact of computer-mediated communication environments on foreign language learning: A review of the literature. *Teaching English with Technology, 14*(2), 67–86.

Mak, B. (2011). An exploration of speaking anxiety with Chinese ESL learners. *TESOL Quarterly, 45*(3), 44–68.

Manda, T. R., & Irawati, L. (2021). The correlation of EFL students' speaking anxiety and their speaking performance. *Eralingua: Jurnal Pendidikan Bahasa Asing dan Sastra, 5*(2), 449–458.

Mandokhail, S., Khan, F. R., & Malghani, M. (2018). Impact of ESL learners' self-esteem on their oral proficiency. *International Journal of English Linguistics, 8*(3), 210–222.

Marzec-Stawiarska, M. (2015). Investigating foreign language speaking anxiety among advanced learners of English. In *Issues in teaching, learning and testing speaking in a second language* (pp. 103–119). Springer.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy, 36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

McCarthy, J. M., & Goffin, R. D. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology, 57*, 607–637.

McGivney, E. (2025). *Leveraging agency in virtual reality to enable situated learning*. arXiv. https://doi.org/10.48550/arXiv.2504.12593

McGivney, M. (2025). Embodied conversational agents as scaffolds for immersive language learning: Situated participation in virtual environments. *Computers & Education, 215*, Article 105046.

Michel, J., & Lehuen, J. (2002, June). Conception of a language learning environment based on the communicative and actional approaches. In *International Conference on Intelligent Tutoring Systems* (pp. 651–660). Springer.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems, E77-D*(12), 1321–1329.

Mishan, F. (2022). Language learning materials in the digital era. In *Materials development for language teaching* (pp. 57–77). Routledge.

Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies, 29*(3), 3195–3217.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309–326.

Moreno, R., Mayer, R., & Lester, J. (2000). Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. In *EdMedia+ innovate learning* (pp. 776–781). Association for the Advancement of Computing in Education (AACE).

Motlagh, F. A., Jafari, A. S., & Yazdani, Z. (2014). A general overview of task-based language teaching (TBLT), from theory to practice. *International Journal of Language and Linguistics, 2*(5-1), 1–11.

Mthembu, B., & Pillay, P. (2014). The development of strategies to teach listening and speaking skills to English Second Language (ESL) learners in African society: Insight from the Pinetown District, KwaZulu Natal. *Assessment, 47*, 48.

Muchnick, A. G., & Wolfe, D. E. (1982). Attitudes and motivations of American students of Spanish. *The Canadian Modern Language Review, 38*(2), 262–281.

Nam, V. H., & Nguyen, T. K. C. (2022). The impact of AI chatbot on customer willingness to pay: An empirical investigation in the hospitality industry. *Journal of Trade Science*.

Naseer, F., Khalid, U., Qammar, M. Z., & Kashif, H. (2024). Chatbots as conversational partners: Their effectiveness in facilitating language acquisition and reducing foreign language anxiety. *Journal of Applied Linguistics and TESOL (JALT), 7*(4), 238–255.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nazligul, M. D., Yilmaz, M., Gulec, U., Yilmaz, A., Isler, V., O'Connor, R. V., & Clarke, P. (2019). Interactive three-dimensional virtual environment to reduce the public speaking anxiety levels of novice software engineers. *IET Software, 13*(2), 152–158.

Newton, J., & Nation, I. S. P. (2020). *Teaching ESL/EFL listening and speaking* (2nd ed.). Routledge.

Nguyen, H. A. (2024). Harnessing AI-based tools for enhancing English speaking proficiency: Impacts, challenges, and long-term engagement. *International Journal of AI in Language Education, 1*(2), 18–29.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence. *Presence: Teleoperators and Virtual Environments, 12*(5), 481–494.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.

Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly, 25*(2), 279–295.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Ochoa, V. C., Lara Tejada, S. V., Chamorro Ortega, C. P., Shagñay Vimos, S. I., & Aucancela Aucanshala, M. de J. (2025). The efficacy of experiential learning in English for Specific Purposes (ESP): A mixed-methods study on speaking skills. *Polo del Conocimiento*.

Ockey, G. J., Chukharev-Hudilainen, E., & Hirch, R. R. (2023). Assessing interactional competence: ICE versus a human partner. *Language Assessment Quarterly*, 1–22.

Ogawa, K. (2025). Development of task-based language teaching in virtual reality: A case study. *Technology in Language Teaching & Learning, 7*(3), 102540–102540.

Ogawa, T. (2025). Exploring the affective dimension of AI-mediated interaction in L2 speaking practice. *Language Learning & Technology, 29*(2), 77–94.

Ogilvie, G., & Dunn, W. (2010). *Taking teacher education to task: Exploring the role of teacher education in promoting the utilization of task-based language teaching* [Unpublished research paper]. University of Alberta.

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*, 114.

OpenAI. (2023). *ChatGPT: Optimizing language models for dialogue* [Large language model]. https://openai.com

Ortega, L., & González-Lloret, M. (2014). *Technology-mediated TBLT*.

Ovsiannikova, U. (2024). *The effect of immersive conversational AI on immigrant students' second language production and willingness to communicate* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context*. Routledge.

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*(3), 1–23. https://doi.org/10.1111/bjop.12290

Panagiotidis, P. (2021). Augmented and mixed reality in language learning. *European Journal of Education, 4*(2), 27–43.

Paneerselvam, A., & Mohamad, M. (2019). Learners' challenges and English educators' approaches in teaching speaking skills in an ESL classroom: A literature review. *Creative Education, 10*(13), 3299–3305.

Pang, Y., He, Y., & Marlow, D. (2025). Employing large language models in college English teaching: Insights from China. *International Journal of Social Sciences, Language and Linguistics*.

Pellettieri, J. (1999). *Why-talk? Investigating the role of task-based interaction through synchronous network-based communication among classroom learners of Spanish* [Unpublished doctoral dissertation]. University of California at Davis.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence in the virtual foreign language classroom. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 59–86). Cambridge University Press.

Peng, J. E., & Woodrow, L. (2010). Willingness to communicate in English: A model in the Chinese EFL classroom context. *Language Learning, 60*(4), 834–876.

Philip, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children's task-based interactions. *Language Learning, 56*(3), 537–575.

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal, 76*(1), 14–26.

Phillips, T., Saleh, A., Glazewski, K. D., Hmelo-Silver, C. E., Mott, B., & Lester, J. C. (2022). Exploring the use of GPT-3 as a tool for evaluating text-based collaborative discourse. *Companion Proceedings of the 12th*, 54.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and instruction. In G. Crookes & S. Gass (Eds.), *Tasks and second language learning* (pp. 9–34). Multilingual Matters.

Pinatih, I. G. A. D. P. (2021). Improving students' speaking skill through role-play technique in 21st century. *Journal of Educational Study, 1*(1), 103–108.

Pontillas, M. S. D. (2020). Reducing the public speaking anxiety of ESL college students through popsispeak. *3L: Language, Linguistics, Literature, 26*(1), 91–105.

Powell, D. M., Stanley, D. J., & Brown, K. N. (2018). Meta-analysis of the relation between interview anxiety and interview performance. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 50*(4), 195.

Pratama, M. R., Sakhiyya, Z., Bahri, S., & Yuliati, Y. (2024). Investigating ESP students and speaking anxiety. *Proceedings of UNNES-TEFLIN National Conference, 6*, 4002.

Qing, X. (2011). Role play – An effective approach to developing overall communicative competence. *Cross-Cultural Communication*.

Rachmawati, D. I. (2020). *EFL learners' speaking anxiety levels and achievement among English department's fifth semester students of Universitas Airlangga* [Doctoral dissertation, Universitas Airlangga].

Rajendran, T., & Yunus, M. M. (2021a). A systematic literature review on the use of mobile-assisted language learning (MALL) for enhancing speaking skills among ESL and EFL learners. *International Journal of Academic Research in Progressive Education and Development, 10*(1), 586–609.

Rajendran, T., & Yunus, M. M. (2021b). Chatterpix Kids: A potential mobile app for helping primary ESL pupils improve their speaking fluency. *International Journal of Learning, Teaching and Educational Research, 20*(4), 18–42.

Ramamuruthy, V. (2019). The effects of task-based approach on speaking anxiety among ESL low proficiency diploma students. *Universal Journal of Educational Research, 7*(6), 1363–1375.

Reynolds, L., & McDonell, K. (2021). *Prompt engineering for AI: Guiding large language models with precision*. arXiv. https://arxiv.org/abs/2102.04074

Richards, J. (2008). *Teaching, listening and speaking from theory to practice*. Cambridge University Press.

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge University Press.

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.

Richardson, J. C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Networks, 7*(1), 68-88.

Rodrigues, I. (2025). Exploring the impact of AI on ESL learning: Students' perspectives and challenges. *INTED2025 Proceedings*, 1739–1745.

Rokhsaritalemi, S., Rezaei, M., & Akbari, H. (2020). Mixed reality applications: A review. *Journal of Ambient Intelligence and Humanized Computing, 11*(8), 3337–3354. https://doi.org/10.1007/s12652-020-01842-0

Rubesch, T. (2013). Interactive competence in student use of a conversational agent. *Studies in Linguistics and Language Teaching, 24*, 157–172.

Rusli, R., Hashim, H. U., Hashim, H., & Yunus, M. M. (2019). Interactive Learning: An Innovation for English Language Acquisition. *International Journal of Innovative Technology and Exploring Engineering, 9*(2), 1187–1190.

Ryan, A. M., Bradburn, J., Bhatia, S., Beals, E., Boyce, A. S., Martin, N., & Conway, J. (2021). In the eye of the beholder: Considering culture in assessing the social desirability of personality. *Journal of Applied Psychology, 106*(3), 452.

Sajjadi, P., Zhao, J., Wallgrün, J. O., Furman, T., La Femina, P. C., Fatemi, A., ... & Klippel, A. (2020, December). The effect of virtual agent gender and embodiment on the experiences and performance of students in virtual field trips. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 221–228). IEEE.

Salaberry, M. R. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal, 85*(1), 39–56.

Salom, S. N. (2023). An investigation on the efficacy of CALL in enhancing comprehension.

Sam, D. P., & Shalini, R. (2021). Limitations and advantages in implementing MALL in the tertiary ESL classrooms: A review. *International Journal of Recent Technology and Engineering, 9*(5), 27–32.

Samani, E., Nordin, N., Mukundan, J., & Samad, A. (2015). Patterns of negotiation of meaning in English as second language learners' interactions. *Advances in Language and Literary Studies, 6*(1), 16–25.

Sarem, S. N., & Shirzadi, Y. (2014). A critical review of the interactionist approach to second language acquisition. *Journal of Applied Linguistics and Language Research, 1*(1), 62–74.

Satar, H. M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL, 25*(1), 122–142.

Satar, H. M., & Özdener, N. (2008). The effects of synchronous computer-mediated communication on speaking anxiety and performance. *Computer Assisted Language Learning, 21*(1), 49–62.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology, 13*(1), 96–120.

Savasci, M. (2013). Why are some students reluctant to use L2 in EFL speaking classes? An action research in tertiary level. *Procedia-Social and Behavioral Sciences, 116*, 2682–2686. https://doi.org/10.1016/j.sbspro.2014.01.635

Savignon, S. J. (1972). *Communicative competence: An experiment in foreign language teaching*. Center for Curriculum Development.

Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice*. Addison-Wesley.

Savignon, S. J. (1991). Communicative language teaching: State of the art. *TESOL Quarterly, 25*(2), 261–277. https://doi.org/10.2307/3587463

Schegloff, E. A., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics, 22*, 3–31.

Schmidt, S. C. E., Gnam, J.-P., Kopf, M., Rathgeber, T., & Woll, A. (2020). The influence of cortisol, flow, and anxiety on performance in e-sports: A field study. *BioMed Research International, 2020*, Article 9651245. https://doi.org/10.1155/2020/9651245

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1–39.

Schroeder, R. C. (2021). *Duolingo and HelloTalk: Analyzing and evaluating language learning applications using SLA theory and user feedback* [Doctoral dissertation, University of Illinois at Urbana-Champaign].

Schuetzler, R. M., Giboney, J. S., Grimes, G. M., & Nunamaker Jr, J. F. (2018). The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems, 114*, 94–102.

Sevim, O. (2014). Effects of drama method on speaking anxieties of preservice teachers and their opinions about the method. *Educational Research and Reviews, 9*(18), 734–742.

Shachter, J., Kangas, M., Sweller, N., & Stewart, J. (2022). Affective states and heart rate response: Measuring foreign language speaking performance reactions in a Japanese university classroom. *Journal for the Psychology of Language Learning, 4*(2), 1–23.

Sharma, D. (2019). *Software development kits: Concepts and best practices*. Packt Publishing.

Shawar, B. A., & Atwell, E. (2007). Fostering language learner autonomy through adaptive conversation tutors. In *Proceedings of the Fourth Corpus Linguistics Conference* (Vol. 3, pp. 186–193).

Shih, Y. C., & Yang, M. T. (2008). A collaborative virtual environment for situated language learning using VEC3D. *Educational Technology & Society, 11*(1), 56–68.

Shikun, S., Grigoryan, G., Huichun, N., & Harutyunyan, H. (2024). AI chatbots: Developing English language proficiency in EFL classroom. *Arab World English Journal (AWEJ), Special Issue on ChatGPT*, 292–305.

Shim, Y. S. (2007). Negotiation of meaning between an L2 teacher and students in face-to-face interactions and CMC. *English Teaching, 62*(3), 265–288.

Sholikhi, F. (2022). Communication students' anxiety in speaking skill practice. *ELT Forum: Journal of English Language Teaching, 11*(1), 31–40. https://doi.org/10.15294/elt.v11i1.49533

Sim, K. N., Yoon, H. J., & Lee, J. S. (2021). Situated task-based learning in virtual worlds: Promoting engagement and intercultural communication. *Journal of Language Teaching and Learning, 11*(3), 45–63.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616.

Slater, M., Brogni, A., & Steed, A. (2004). Physiological responses to breaks in presence: A pilot study. *Presence: Teleoperators and Virtual Environments, 13*(4), 336–342. https://doi.org/10.1162/1054746041944803

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., ... & Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments, 15*(5), 553–569. https://doi.org/10.1162/pres.15.5.553

Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal, 87*(1), 38–57.

Smith, B. (2009). The relationship between scrolling, negotiation, and self-initiated self-repair in a SCMC environment. *CALICO Journal, 26*, 231–245.

Smith, M. D., & McCurrach, D. P. (2024). Development of task-based language teaching in virtual reality: A case study. *Technology and Language Learning, 7*(3).

Smith, M. J., Ginger, E. J., Wright, K., Wright, M. A., Taylor, J. L., Humm, L. B., ... & Fleming, M. F. (2014). Virtual reality job interview training in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*, 2450–2463.

Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment, 21*(3), 425.

Speicher, M., Hall, B. D., & Nebeling, M. (2019, May). What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).

Sreena, S., & Ilankumaran, M. (2018). Developing productive skills through receptive skills–a cognitive approach. *International Journal of Engineering & Technology, 7*(4.36), 669–673.

Srinivasan, M. A., & Basdogan, C. (1997). Haptics in virtual environments and teleoperation. *Presence: Teleoperators and Virtual Environments, 6*(4), 389–421.

Stapleton, P. (2020). [Chapter title]. In C. L. Fracasso, S. Krippner, & H. L. Friedman (Eds.), *Holistic treatment in mental health: A handbook practitioner's perspectives* (pp. 136–151). McFarland.

Stockwell, G., & Hubbard, P. (2013). *Some emerging principles for mobile-assisted language learning*. Monterey Institute of International Studies.

Stretton, T., Cochrane, T., & Narayan, V. (2018). Exploring mobile mixed reality in healthcare higher education: A systematic review. *Research in Learning Technology, 26*.

Stupar-Rutenfrans, S., Ketelaars, L. E., & van Gisbergen, M. S. (2017). Beat the fear of public speaking: Mobile 360 video virtual reality exposure training in home environment reduces public speaking anxiety. *Cyberpsychology, Behavior, and Social Networking, 20*(10), 624–633.

Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning, 71*(4), 1149–1193.

Sun, R., et al. (2023). *Collaborative system design of mixed reality communication for medical training*. arXiv.

Tafazoli, D., & Gómez-Parra, M. E. (2017). Robot-assisted language learning: Artificial intelligence in second language acquisition. *Current and Future Developments in Artificial Intelligence, 1*, 370–396. https://doi.org/10.2174/9781681085029117010015

Tai, T. Y., & Chen, H. H. J. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research, 59*(7), 1272–1293.

Takimoto, M. (2014). Evaluating the effects of input-based approaches to the teaching of pragmalinguistics and sociopragmatics in second language pragmatics: A case of English request hedges. *IAFOR Journal of Language Learning, 1*(1), Article n1.

Tanaka, H., Matsuda, T., & Nakamura, S. (2021). AI chatbots as speaking partners: An empirical study. *ReCALL, 33*(3), 256–271. https://doi.org/10.1017/S0958344021000140

Tanveer, M. (2007). *Investigation of the factors that cause language anxiety for ESL/EFL learners in learning speaking skills and the influence it casts on communication in the target language* [Doctoral dissertation, University of Glasgow].

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *Applied Linguistics, 30*(4), 473–494.

Tee, X. T., Joanna, T. A. T., & Kamarulzaman, W. (2020). A systematic review of self-coping strategies used by university students to cope with public speaking anxiety. *English Language Teaching, 13*(10), 57–73.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*(4), 525–547. https://doi.org/10.1177/0959354394044004

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis–Erratum. *Studies in Second Language Acquisition, 41*(2), 489.

Thao, T. Q., & Nguyet, D. T. N. (2019). Four aspects of English speaking difficulties encountered by tertiary English-majored students. *Ho Chi Minh City Open University Journal of Science-Social Sciences, 9*(2), 53–64.

Togimin, N., & Jaafar, H. (2020). Role play activities: An innovative approach to teaching speaking skills in an ESL classroom. *LSP International Journal, 7*(2), 19–36.

Toyoda, E., & Harrison, R. (2002). Categorization of text chat communication between learners and native speakers of Japanese. *Language Learning & Technology, 6*, 82–99.

Tsiplakides, I., & Keramida, A. (2009). Helping students overcome foreign language speaking anxiety in the English classroom: Theoretical issues and practical recommendations. *International Education Studies, 2*(4), 39–44.

Tu, C. H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education, 16*(3), 131–150.

Tuan, N. H., & Mai, T. N. (2015). Factors affecting students' speaking performance at Le Thanh Hien High School. *Asian Journal of Educational Research, 3*, 8–23.

Tudini, V. (2003). Using native speakers in chat. *Language Learning & Technology, 7*, 141–159.

Uwizeyimana, V. (2018). *The effect of mobile-assisted language learning on Rwandan university students' speaking proficiency* [Unpublished dissertation].

Van den Branden, K. (2006). *Task-based language education from theory to practice*. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://arxiv.org/abs/1706.03762

Vazquez, C. D., Nyati, A. A., Luh, A., Fu, M., Aikawa, T., & Maes, P. (2017, May). Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2172–2179).

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System, 127*, Article 103533.

Wang, H., Pi, Z., & Hu, W. (2019). The instructor's gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning, 35*(1), 42–50. https://doi.org/10.1111/jcal.12309

Wang, J., Yang, H., Shao, R., Abdullah, S., & Sundar, S. S. (2020). Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Conference on Human Factors in Computing Systems - Proceedings*.

Wang, X., Laffey, J., Xing, W., Ma, Y., & Stichter, J. (2016). Exploring embodied social presence of youth with autism in a 3D collaborative virtual learning environment: A case study. *Computers in Human Behavior, 55*, 310–321. https://doi.org/10.1016/j.chb.2015.09.006

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 humanised agent from the community of inquiry. *Computer Assisted Language Learning, 37*(4), 814–840.

Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL, 18*, 122–146.

Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual Immersive Language Learning and Gaming Environment: Immersion and presence. *British Journal of Educational Technology, 48*(2), 431–450.

Wang, Y., Grist, M., & Grant, S. (2021). Into the real world: Autonomous and integrated Chinese language learning through a 3D immersive experience. In *Contextual language learning: Real language learning on the continuum from virtuality to reality* (pp. 119–145). Springer.

Warschauer, M. (1996). *Computer-assisted language learning: An introduction*.

Weiser, M. (1999). The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review, 3*(3), 3–11.

Weng, C., Rathinasabapathi, A., Weng, A., & Zagita, C. (2019). Mixed reality in science education as a learning support: A revitalized science book. *Journal of Educational Computing Research, 57*(3), 777–807.

Westhoff, J. G. (2009). *A priori assessment of language learning tasks by practitioners* [Unpublished research paper]. Utrecht University.

Whitley-Walters, D., & Muhammad, J. (2021, January). The analysis of implementing augmented reality within remote learning. In *ADMI 2021: The Symposium of Computing at Minority Institutions*.

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand, 14*, Article 100224.

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636.

Winkler, R., & Söllner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Annual Meeting Proceedings, 2018*(1), 1–40.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal, 37*(3), 308–328.

Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2024). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning, 4*(2), 300–325.

Xing, C., & Saeed, M. A. (2025). A systematic review on artificial intelligence (AI) technologies in ESL/EFL speaking skills. *International Journal of TESOL Studies*, Article 250908. https://doi.org/10.58304/ijts.250908

Xin-ming, L. (2010). *Learners' needs analysis in task-based language teaching* [Unpublished research paper]. Zhejiang Gongshang University.

Yamada, M., & Akahori, K. (2007). Social presence in synchronous CMC-based language learning: How does it affect the productive performance and consciousness of learning objectives? *Computer Assisted Language Learning, 20*(1), 37–65. https://doi.org/10.1080/09588220601118503

Yan, W., Lowell, V. L., & Yang, L. (2024). Developing English language learners' speaking skills through applying a situated learning approach in VR-enhanced learning experiences. *Virtual Reality, 28*(4), 167.

Yanguas, Í. (2010). Oral computer-mediated interaction between L2 learners: It's about time. *Language Learning & Technology, 14*, 72–93.

Yanguas, Í. (2012). Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO Journal, 29*, 507–531.

Yao, C., Kanjanakate, S., & Jantharajit, N. (2024). Bridging language learning and cognitive growth: The power of situated learning and task-based language teaching in modern education. *JET (Journal of English Teaching), 10*(3), 320–334.

Yaoqing, Z. (2021, December). Review of enlightenment of emotional factors in Krashen's affective filter hypothesis on English teaching research in college. In *2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021)* (pp. 1518–1524). Atlantis Press.

Yin, Q., & Satar, M. (2020). English as a foreign language learner interactions with chatbots: Negotiation for meaning. *International Online Journal of Education and Teaching, 7*(2), 390–410.

Ying, Y. H., Siang, W. E. W., & Mohamad, M. (2021). The challenges of learning English skills and the integration of social media and video conferencing tools to help ESL learners coping with the challenges during COVID-19 pandemic: A literature review. *Creative Education, 12*(7), 1503–1516.

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(4), 426–439.

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. *Educational Technology Research and Development, 73*(2), 815–874.

Zacarin, M. R. J., Borloti, E., & Haydu, V. B. (2019). Behavioral therapy and virtual reality exposure for public speaking anxiety. *Trends in Psychology, 27*(2), 491–507.

Zakaria, N., Rahmat, N. H., Aripin, N., Jasman, N. H., & Ibrahim, N. (2019). Exploring perceptions of role play activities in English as a second language (ESL) classrooms. *International Journal of Engineering and Advanced Technology, 8*(5), 868–872.

Zare, P., & Othman, M. (2015). Students' perceptions toward using classroom debate to develop critical thinking and oral communication ability. *Asian Social Science, 11*(9), 158.

Zechner, O., Schrom-Feiertag, H., Wespi, R., Pretolesi, D., Nguyen, Q., & Tscheligi, M. (2024). Enhancing mixed reality simulation training technology with real-time performance visualization: Mixed methods study with medical first responders. *JMIR XR and Spatial Computing, 1*(1), e57655. https://doi.org/10.2196/57655

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence, 4*, 100134.

Zhan, T., Yin, K., Xiong, J., He, Z., & Wu, S. T. (2020). Augmented reality and virtual reality displays: perspectives and challenges. *Iscience, 23*(8).

Zhang, R., Zou, D., & Cheng, G. (2024). Concepts, affordances, and theoretical frameworks of mixed reality enhanced language learning. *Interactive Learning Environments, 32*(7), 3624–3637.

Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal, 103*(4), 763–781.

Zheng, D., Li, N., & Zhao, Y. (2008, March). Learning Chinese in Second Life Chinese language school. In *CALICO Annual Conference, San Francisco, CA*.

Zhou, W. (2023). *Chat GPT integrated with voice assistant as learning oral chat-based constructive communication to improve communicative competence for EFL earners*. arXiv preprint arXiv:2311.00718.

Ziegler, N. (2016). Taking technology to task: Technology-mediated TBLT, performance, and production. *Annual Review of Applied Linguistics, 36*, 136–163.

Appendix A

Mock Job Interview Pretest and Post Test

Context:

Imagine you're applying for your dream job. This position aligns perfectly with your skills, interests, and long-term career goals. The company is known for its inclusive culture and values teamwork, innovation, and personal growth. Today, you have an interview with the hiring manager to discuss your fit for the role.

The examiner will welcome the participant and will ask them about their dream job. This same dream job will be used in the posttest as well. This is a roleplay activity designed as a pretest and post test. Students will be informed about that.

Introduction:

"Good morning/afternoon! Thank you for coming in today. I'm [Interviewer's Name], and I'll be conducting your interview. We're excited to learn more about you and see how you might fit with our team. This should take 20 minutes of your time. Before we begin, do you have any questions, or is there anything you'd like to discuss?"

Standardized Interview questions:

1. Tell me about yourself.
2. Walk me through your resume.
3. Why should we hire you?
4. What can you bring to the company?
5. What are your greatest strengths?
6. What do you consider to be your weaknesses?
7. What is your greatest professional achievement?
8. Tell me about a challenge or conflict you've faced at work, and how you dealt with it.

Appendix B

Mock Job Interview Rubric

Job Interview Speaking Rubric (1–3 Scale)

Criteria	**3 – Excellent**	**2 – Satisfactory**	**1 – Needs Improvement**
Fluency	Speaks smoothly and confidently with natural pacing and minimal hesitation.	Generally fluent with occasional pauses or hesitations that slightly affect flow.	Frequent pauses and hesitations disrupt communication.
Pronunciation	Pronunciation is clear and accurate; sounds are articulated well, enhancing understanding.	Pronunciation is mostly clear and understandable with some minor errors.	Pronunciation is often unclear, leading to difficulty in understanding.
Intonation & Stress	Uses natural intonation and correct stress patterns that support meaning and engagement.	Generally uses appropriate intonation and stress, though inconsistently.	Monotone or incorrect stress patterns make speech sound unnatural or confusing.
Grammar & Sentence Structure	Uses a range of sentence structures accurately; grammar is consistently correct.	Uses mostly correct basic structures; occasional errors in complex sentences.	Frequent grammatical errors and limited sentence variety make understanding difficult.
Vocabulary Use	Uses a wide and relevant range of vocabulary accurately, including job-specific terms.	Uses appropriate vocabulary with some repetition or minor inaccuracies.	Limited vocabulary with frequent inaccuracies; lacks job-specific terminology.

Content & Relevance of Answers	Provides detailed, relevant, and well-prepared answers showing strong understanding of the job.	Answers are clear and mostly relevant, showing general preparation and understanding.	Answers are vague, off-topic, or show limited preparation and understanding of the job.
Use of Work-Based Examples	Provides specific, relevant examples from work or experience that strongly support responses.	Provides some relevant examples but lacks detail or clear connection to responses.	Provides few or no examples, or examples that are irrelevant or unclear.

Notes for Use:

- Total Points: The maximum score is 21 points (7 criteria × 3 points each)
- Scoring: Allocate points based on the performance level in each category (1 = Needs Improvement, 2 = Satisfactory, 3 = Excellent). Add the points to calculate the total score.

Appendix C

Foreign Language Classroom Anxiety Scale (FLCAS) Pretest and Posttest

1. I never feel quite sure of myself when I am speaking in my foreign language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

2. I don't worry about making mistakes in language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

3. I tremble when I know that I'm going to be called on in language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

4. It frightens me when I don't understand what the teacher is saying in the foreign language.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

5. It wouldn't bother me at all to take more foreign language classes.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

6. During language class, I find myself thinking about things that have nothing to do with the course.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

7. I keep thinking that the other students are better at languages than I am.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

8. I am usually at ease during tests in my language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

9. I start to panic when I have to speak without preparation in language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

10. I worry about the consequences of failing my foreign language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

11. I don't understand why some people get so upset over foreign language classes.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

12. In language class, I can get so nervous I forget things I know.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

13. It embarrasses me to volunteer answers in my language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

14. I would not be nervous speaking the foreign language with native speakers.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

15. I get upset when I don't understand what the teacher is correcting.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

16. Even if I am well prepared for language class, I feel anxious about it.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

17. I often feel like not going to my language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

18. I feel confident when I speak in foreign language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

19. I am afraid that my language teacher is ready to correct every mistake I make.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

20. I can feel my heart pounding when I'm going to be called on in language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

21. The more I study for a language test, the more confused I get.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

22. I don't feel pressure to prepare very well for language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

23. I always feel that the other students speak the foreign language better than I do.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

24. I feel very self-conscious about speaking the foreign language in front of other students.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

25. Language class moves so quickly I worry about getting left behind.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

26. I feel more tense and nervous in my language class than in my other classes.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

27. I get nervous and confused when I am speaking in my language class.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

28. When I'm on my way to language class, I feel very sure and relaxed.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

29. I get nervous when I don't understand every word the language teacher says.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

30. I feel overwhelmed by the number of rules you have to learn to speak a foreign language.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

31. I am afraid that the other students will laugh at me when I speak the foreign language.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

32. I would probably feel comfortable around native speakers of the foreign language.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

33. I get nervous when the language teacher asks questions which I haven't prepared in advance.

Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree

Appendix D

Reflection Interview Questions

Protocol for Interview:

Welcome, and thank you for volunteering to participate in this study.

Today, our discussion will focus on your experiences with practicing with [insert condition] for job interview skills. I aim here to interview you to understand if this method has been effective in enhancing your speaking skills. I will be asking you a series of questions related to this topic. If you prefer, I can provide you with these questions in writing.

Please note that this session will be video recorded to aid in my research analysis. However, your participation is entirely voluntary, and you may withdraw from the interview at any point should you wish to do so.

It's important to emphasize that there are no right or wrong answers in this conversation. Your honest and candid responses are what matter most. Additionally, you are encouraged to share any questions or thoughts you might have at any time during our discussion.

Before we begin, do you have any questions or need any clarifications?

Introduction to Practice Condition:

- Could you briefly describe your initial impressions when you were introduced to your assigned practice condition (Mixed Reality, ChatGPT Voice App, or Traditional Techniques)?

Experience with the Practice Condition:

- How did the practice sessions under your assigned condition make you feel about your speaking abilities in English during job interviews?
- Can you share a specific example or moment during the practice sessions that significantly impacted your confidence or anxiety levels?

Effectiveness of the Practice Condition:

- In what ways do you feel the practice condition was effective in improving your speaking performance for job interviews?
- Were there any aspects of the practice condition that you found particularly challenging or unhelpful? Please elaborate.

Comparison with Other Conditions (Hypothetical for participants who did not experience them):

- Based on what you know or can imagine, how do you think your experiences might have differed if you you used (e.g chatGPT, human instructor, immersive application)?

Impact on Anxiety and Performance:

- How do you believe your assigned practice condition affected your anxiety levels related to job interviews in English?
- Do you feel there was a noticeable improvement in your job interview performance after participating in the practice sessions? Can you provide an example?

Recommendations and Suggestions:

- What improvements or changes would you suggest for your practice condition to better support non-native English speakers in job interview preparations?
- Would you recommend your assigned practice condition to other non-native English speakers preparing for job interviews? Why or why not?

General Feedback:

- Is there anything else about your experience with the practice condition that you'd like to share or any additional support you wish you had received?

Closing:

- Thank you