Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Human-Guided AI Interaction in Task-Based Language Teaching: Effects of ChatGPT and Embodied Multimodal Conversational AI in Mixed Reality on ESL Learners' Job Interview Anxiety and Oral Performance

**Amany Alkhayat**

amanyudemy@gmail.com

Media University of Applied Sciences

**Research Article**

**Keywords:** ESL speaking performance, foreign language anxiety, mixed reality, conversational AI, English for Specific Purposes, job interview, FLCAS

**Posted Date:** August 12th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9875876/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Abstract

This quasi-experimental mixed-methods study investigated the effects of two AI-mediated conversational modalities on ESL learners' oral performance and speaking anxiety in a high-stakes English for Specific Purposes (ESP) context. Seventy-five adult ESL learners were assigned to one of three conditions: traditional instructor-led instruction (control), ChatGPT voice mode, or a Multimodal Embodied Conversational AI agent (MECAI) in Mixed Reality (MR). Over a six-week intervention, participants completed pre- and post-intervention mock job interviews assessed for fluency, vocabulary appropriateness, and grammatical accuracy. Speaking anxiety was measured using the Foreign Language Classroom Anxiety Scale (FLCAS) and physiological heart rate monitoring. Two-way repeated measures ANOVA revealed significant improvements in oral performance across all groups, with the MECAI group demonstrating the greatest gains (pre: 9.36; post: 18.52), followed by ChatGPT (pre: 8.96; post: 15.52) and the control group (pre: 7.92; post: 12.88). Pairwise comparisons confirmed statistically significant differences between all three conditions ($p < 0.001$). Although descriptive trends suggested a pattern of anxiety reduction in the MECAI group across both FLCAS and heart rate, neither anxiety measure reached statistical significance. Qualitative thematic analysis of structured interviews identified four themes: AI's role in shaping oral confidence, immersion and sense of presence in the MR environment, the indispensable role of the human instructor, and learner strategies for managing language anxiety. Findings suggest that embodied conversational AI in mixed reality offers meaningful advantages for ESP speaking development, while underscoring the continued centrality of human instruction in technology-enhanced language learning.


# 1 Introduction

Speaking in a second or foreign language is considered an important component of communicative competence for ESL/EFL learners, particularly in academic and professional contexts where effective spoken communication is essential for shared meaning construction and mutual understanding (Huang et al., 2021). In high-stakes speaking tasks within an increasingly globalized labor market, ESL learners need to practice language differently from general English courses to transfer skills for employability, career mobility, and access to work-related opportunities (Zhang, 2019). Tasks such as job interviews require learners to demonstrate linguistic accuracy, fluency, pragmatic appropriateness, and confidence under evaluative pressure (Ghafar, 2022; Rao, 2019). However, barriers such as large classroom sizes, lack of personalization, and limited opportunities for authentic interaction restrict speaking development and contribute to affective challenges (Brown, 2001; Fryer & Carpenter, 2006).

Among these challenges, speaking anxiety has been consistently identified as a critical impediment to oral performance and is conceptualized as a situation-specific form of anxiety in language learning contexts (Horwitz et al., 1986; MacIntyre & Gardner, 1991). Anxiety is triggered by fear of negative self-evaluation, peer and teacher judgment, communication apprehension, and test-related stress (Cantos et al., 2024; Tania et al., 2025; Alkhayat, 2017). Empirical findings show that heightened anxiety exhausts cognitive resources necessary for speech production, leading to mental blocks, impaired performance,

reduced lexical access, grammatical inaccuracies, and avoidance behaviors (Eysenck et al., 2007; Abbasi et al., 2019; Botes et al., 2020; Teimouri et al., 2019). Moreover, In ESP job-interview preparation, curriculum and instructor limitations can make it difficult to address learners' diverse career goals. Recent ESP research emphasizes that effective instruction requires needs-based, discipline-specific materials; however, many programs still struggle to align classroom content with students' professional communication demands (Zeng & Della, 2024; Dülger & Ugurlu, 2025). This creates challenges for instructors who may be expected to prepare students for varied workplace interviews without having specialized knowledge of each learner's target field (Hyland, 2006). These challenges become especially evident in job interview preparation, where learners from diverse professional backgrounds often require specialized vocabulary, industry-specific knowledge, and tailored communicative practice that general ESP curricula may not adequately address.

To address these challenges, research in general ESL/EFL contexts has explored diverse types of interventions aimed at improving speaking performance while reducing anxiety. Pedagogically, learner-centered approaches and low-anxiety classroom environments have been shown to encourage participation and reduce fear of evaluation (Young, 1991). Drama-based techniques and communicative activities further support engagement and lower anxiety when appropriately scaffolded (Sevim, 2014). However, performance-oriented activities such as oral presentations may increase anxiety due to heightened expectations and evaluative pressure (Altunkaya, 2018). These findings indicate that while communicative approaches are beneficial, their effectiveness depends on alignment with learners' psychological readiness and curriculum alignment.

When it comes to research on ESP, this area is way underdeveloped. There's a few research that shows several limitations. Alwi et al. (2025) focused on VR-enhanced task-based learning for interview anxiety, but the study appears limited to one intervention type, making it difficult to know whether VR is more effective than other ESP interview-preparation methods when it comes to more diverse populations where instructors only have expertise in second language instruction. Takaya (2024) examined learning strategies for job-interview skill development, but strategy-focused work may not fully address curriculum limitations, instructor expertise, or the need for discipline-specific interview practice. Fareen (2017) developed an employment-oriented ESP course for IT students, but its focus on one professional field limits its application to ESP classrooms with learners from diverse career backgrounds.

A critical gap in the literature lies in the lack of integrated research examining speaking performance, anxiety, and interaction modality within AI-mediated environments. Additionally, many studies rely primarily on self-reported measures of anxiety, with limited use of physiological indicators such as heart rate, which can provide deeper insights into learners' affective states (Kaplan-Rakowski & Gruber, 2023). Addressing these limitations requires a multimodal and mixed-methods approach that captures both performance and psychological dimensions.

Accordingly, this study investigates the impact of multimodal conversational AI on ESL learners' speaking performance and anxiety in mock job interview contexts. It compares interaction with an

embodied AI agent in Mixed Reality, a voice-based AI system, and traditional instructional methods. The study is guided by the following research questions:

1. How do human-guided traditional instruction, ChatGPT voice interaction, and embodied multimodal conversational AI in mixed reality affect ESL learners' oral performance in task-based mock job interviews from pre- to post-intervention?
2. How do human-guided traditional instruction, ChatGPT voice interaction, and embodied multimodal conversational AI in mixed reality affect ESL learners' job interview anxiety, as measured through self-reported anxiety and physiological indicators such as heart rate?
3. How do ESL learners perceive human-guided traditional instruction, ChatGPT voice interaction, and embodied multimodal conversational AI in mixed reality for task-based job interview preparation?

# 2 Literature Review

Most research in the literature focuses on interventions related to general ESL/EFL contexts. In this area, research has explored a range of interventions targeting both anxiety reduction and speaking performance. Psychological interventions, including Cognitive Behavioral Therapy (CBT), mindfulness, and emotional regulation strategies, have shown effectiveness in reducing anxiety by addressing cognitive distortions, maladaptive thought patterns, and physiological responses (Dincer et al., 2020; Elahemer & Said, 2022; Cahya & Churnia, 2026). However, as these interventions do not provide enough opportunities for speaking practice, their effect is often indirect. Pedagogical interventions have shown greater effectiveness in simultaneously reducing anxiety and improving performance. Researchers employed pedagogical approaches such as learner-centered approaches, collaborative learning, and task-based instruction which promoted gradual exposure and errors normalization (Sotomayor Cantos et al., 2024). Dynamic assessment further enhances these outcomes by integrating scaffolding and feedback during interaction (Anton, 2009; Kayhan, 2025). Technological interventions enhance these approaches by offering alternative, anxiety-free environments for practice. Early forms of Computer-Mediated Communication (CMC) and online interaction facilitated low-pressure possibilities for language use, especially when learners could control timing and edit responses in text-based asynchronous or quasi-synchronous communications (Hubbard, 2008). However, research showed that synchronous or high-exposure formats, such as video recording or real-time conferencing, sometimes increased anxiety due to heightened self-awareness and perceived evaluation (Kaplan-Rakowski & Gruber, 2023). These findings emphasize that technology alone does not guarantee effectiveness; rather, outcomes depend on design, interactivity, and alignment with pedagogical goals.

Within this technological trajectory, Dialogue-based CALL has emerged as a promising intervention that has potential for enhancing oral proficiency. Early rule-based chatbots that relied on pattern matching and predefined scripts, rather than neural networks, provided structured, low-pressure learning environments that enhanced pronunciation, grammatical accuracy, and basic speaking practice (Aliakbari et al., 2025; Huang et al., 2021). Their predictability reduced L2 learners' anxiety and enabled repeated practice, leading to measurable improvements in speaking performance. However, because

these systems lacked contextual awareness, could not handle spontaneous input, and were difficult to align with the curriculum, they felt less authentic, and failed to provide effective communication (Xing & Saeed, 2025).

On the other hand, the development of generative AI chatbots represents a significant advancement over rule-based systems. Powered by large language models, these systems enable context-sensitive, adaptive, and extended dialogue, allowing learners to engage in more authentic and dynamic interactions (Toncelli & Kostka, 2024; Ulla & Teng, 2024; Xing & Saeed, 2025). Research suggests that generative AI can simultaneously improve speaking performance and reduce anxiety by providing personalized feedback and low-pressure interaction (Fathi et al., 2024; Xing & Saeed, 2025). Nevertheless, limitations remain, including potential inaccuracies, lack of pedagogical alignment, and the risk of learner over-reliance, which may reduce deeper cognitive engagement (Spatola, 2024; Wang et al., 2025; Xing & Saeed, 2025).

Additionally, immersive technologies such as Virtual Reality (VR) and Mixed Reality (MR) have introduced new opportunities for language learning through embodied interaction and enhanced social presence. Research shows that VR environments can create simulated contexts that allow learners to practice speaking while benefiting from a degree of psychological distance through avatars, which can reduce anxiety and increase willingness to communicate (Kaplan-Rakowski & Gruber, 2023). MR extends these affordances by integrating digital and physical environments, enabling more contextualized and interactive learning experiences. The inclusion of embodied AI agents in such environments further enhances authenticity by incorporating non-verbal cues and spatial interaction, which are critical for pragmatic competence (Kreijns et al., 2022).

However, the effectiveness of immersive technologies remains mixed. While some studies report reduced anxiety and improved engagement, others highlight challenges such as technological complexity, limited usability, and inconsistent learning outcomes. Moreover, existing research is often limited by small sample sizes, short intervention durations, and a reliance on self-reported measures of anxiety, which may not fully capture real-time emotional responses (Tania et al., 2025; Kaplan-Rakowski & Gruber, 2023).

In sum, the literature suggests that interventions are most effective when they integrate affective support with meaningful communicative practice. Psychological and pedagogical approaches provide foundational support, while technological interventions—particularly generative AI and immersive environments—offer scalable and flexible opportunities for practice. However, no single intervention is universally effective. Rule-based chatbots and poorly designed technological tools often fail due to limited interactivity, while immersive technologies and generative AI show promise but require careful design and further empirical validation. Crucially, there remains a lack of integrated research examining how different AI-mediated interaction modalities influence both speaking performance and anxiety simultaneously, particularly in high-stakes ESP contexts.

# 3 Theoretical Framework

Foreign language anxiety (FLA) has been most systematically conceptualized through the Foreign Language Classroom Anxiety Scale (FLCAS), which frames anxiety as a situation-specific construct arising from the cognitive and social demands of language learning (Horwitz et al., 1986; Xiong et al., 2024; Cahya & Churnia, 2026). Within this framework, communication apprehension, fear of negative evaluation, and test anxiety collectively shape learners' emotional responses to speaking tasks (Tania et al., 2025; Sotomayor Cantos et al., 2024). Empirical research consistently shows that elevated anxiety disrupts attentional control and working memory, leading to reduced fluency, increased pauses, limited lexical access, and greater grammatical inaccuracies (Eysenck et al., 2007; Botes et al., 2020; Teimouri et al., 2019). As a result, learners often experience mental blocks and avoidance behaviors, reinforcing a cyclical relationship between anxiety and poor performance (Abbasi et al., 2019).

While the FLCAS framework has been widely used to evaluate interventions, demonstrating measurable reductions in anxiety across pedagogical and technological contexts (Elahemer & Said, 2022; Kayhan, 2025), it also presents methodological limitations. Its reliance on self-reported data introduces potential bias and limits its ability to capture real-time fluctuations in anxiety (Cahya & Churnia, 2026). Consequently, recent research has advocated for integrating physiological measures, such as heart rate and electrodermal activity, to provide a more objective and dynamic understanding of learners' affective states (Tania et al., 2025; Kaplan-Rakowski & Gruber, 2023). This shift toward mixed-method approaches reflects a broader need to move beyond static assessments of anxiety toward more comprehensive models of speaking performance.

Building on this affective foundation, social constructivist theory offers a complementary perspective by emphasizing the role of interaction in language development. Vygotsky's (1978) concept of the Zone of Proximal Development (ZPD) posits that learning occurs through guided interaction with a More Knowledgeable Other (MKO), who provides scaffolding that enables learners to achieve higher levels of competence (Yu et al., 2024). In ESL/EFL speaking contexts, this scaffolding is critical for supporting learners as they navigate increasingly complex communicative tasks. The integration of AI into language learning aligns with this framework, as conversational agents can function as digital MKOs, offering continuous, adaptive support.

Empirical studies indicate that AI-mediated interaction can enhance speaking performance by providing immediate feedback, personalized scaffolding, and unlimited opportunities for practice (Huang et al., 2021; Aliakbari et al., 2025). Additionally, learners interacting with AI systems often demonstrate increased willingness to communicate and reduced anxiety due to the low-pressure, non-judgmental nature of these environments (Fathi et al., 2024). However, the application of social constructivism in AI contexts is not without limitations. Unlike human interlocutors, AI systems lack genuine intentionality and may provide inaccurate or pedagogically misaligned responses, potentially hindering learning (Spatola, 2024; Wang et al., 2025). Furthermore, over-reliance on AI scaffolding may reduce learner autonomy, as constant access to feedback can limit the development of independent problem-solving

strategies. These concerns highlight the need to critically examine not only the presence of scaffolding but also its quality and long-term effects.

Extending beyond interactional scaffolding, social presence theory provides further insight into how the modality of interaction influences both cognitive and affective outcomes. Social presence refers to the extent to which learners perceive their communication partner as “real” within a mediated environment (Short et al., 1976; Kreijns et al., 2022). Higher levels of social presence are associated with increased engagement, motivation, and perceived learning (Richardson et al., 2017). In language learning, social presence is particularly important because it enables access to non-verbal cues such as gestures, facial expressions, and tone of voice, which support pragmatic competence and interactional fluency (Yamada & Akahori, 2007).

In AI-mediated environments, social presence can be enhanced through human-like voices, avatars, and embodied agents. According to Social Agency Theory, these features encourage learners to treat interactions with AI as social experiences, thereby increasing cognitive engagement (Mayer, 2014). Empirical findings suggest that AI systems with higher social presence can reduce anxiety and improve learning outcomes by fostering emotional connection and more natural communication (Kim et al., 2020; Guo et al., 2023). This progression from text-based to more immersive modalities reflects an important shift in how technology supports language learning.

At the highest level of social presence, immersive technologies such as Virtual Reality (VR) and Mixed Reality (MR) enable embodied interaction within three-dimensional environments. These technologies create a strong sense of 'being there,' allowing learners to engage in realistic communicative scenarios while maintaining a degree of psychological safety (Kreijns et al., 2022). Research indicates that such environments can reduce speaking anxiety by balancing authenticity with reduced fear of direct social judgment (Kaplan-Rakowski & Gruber, 2023). Moreover, the integration of embodied AI agents within these environments enhances both linguistic and pragmatic development by incorporating verbal and non-verbal communication cues.

Despite these advantages, the relationship between social presence and learning outcomes remains complex. Empirical findings are inconsistent, with some studies reporting significant improvements and others showing limited or indirect effects (Xu et al., 2026). Additionally, conceptual ambiguity persists, as social presence is often conflated with related constructs, leading to inconsistencies in measurement and interpretation (Kreijns et al., 2022). Importantly, there is a lack of comparative research examining how different interaction modalities—text-based, voice-based, and embodied—affect both speaking performance and anxiety. Furthermore, few studies integrate social presence with physiological measures, limiting understanding of how perceived 'realness' influences real-time emotional responses.

Task-Based Language Teaching (TBLT) is not a language learning theory but rather a pedagogical approach to language instruction. Rooted in communicative language teaching, TBLT emphasizes the use of meaningful tasks that require learners to use language to achieve specific goals in authentic or real-world contexts. Instead of focusing primarily on the explicit teaching of grammatical forms, TBLT

promotes language acquisition through task completion, interaction, problem-solving, and reflection. A typical TBLT lesson consists of three phases: a pre-task phase that prepares learners for the task, a task cycle in which learners perform communicative activities, and a post-task phase that focuses on feedback, reflection, and language development. Systematic literature reviews generally report that TBLT improves speaking performance, communicative competence, learner confidence, and classroom engagement, especially when tasks are authentic, goal-oriented, and followed by feedback. Reviews also suggest that TBLT can reduce speaking anxiety because learners focus on completing meaningful tasks rather than only producing correct grammar. For EFL speaking, systematic reviews highlight gains in fluency, interaction, vocabulary use, motivation, and confidence, although outcomes depend on task design, teacher guidance, learner proficiency, and assessment quality.

In summary, the literature shows a progression from affective frameworks such as FLCAS, to interactional theories like social constructivism, and finally to modality-focused perspectives such as social presence theory. Together, these frameworks highlight that speaking performance and anxiety are shaped not only by internal cognitive and emotional factors but also by the nature and quality of interaction. However, significant gaps remain in understanding how these dimensions intersect within multimodal AI environments, particularly in high-stakes speaking contexts.

# 4 Methodology

This study employed a quasi-experimental mixed-methods design to examine the effects of multimodal conversational artificial intelligence (AI) on ESL learners' speaking performance and speaking anxiety in a high-stakes English for Specific Purposes (ESP) context. The study compares three instructional conditions across pre- and post-intervention measures: Group 1 (traditional instruction), Group 2 (ChatGPT voice mode), and Group 3 (multimodal embodied conversational AI [MECAI] in Mixed Reality). The intervention was conducted over a six-week period, triangulated quantitative and qualitative data and physiological measures to provide a comprehensive analysis. Students completed pre- and post-tests administered by external examiners, who participated in the study for a fee paid by the researcher. During the mock job interviews, which consisted of eight questions, students completed the FLCAS questionnaire and had their heart rate measured using the Polar OH1 after the pre-test. They then participated in 20 minutes of weekly in-class practice using a TBLT approach before completing a similar post-test (Fig. 1).

Because participants were assigned using a quasi-experimental procedure, baseline equivalence was examined across the three groups prior to outcome interpretation. A one-way ANOVA revealed no significant group differences in pretest baseline 2-minute heart rate, $F(2,72) = 1.18$, $p = .312$, $\eta^2 = .032$, or pretest interview heart rate, $F(2,72) = 1.28$, $p = .283$, $\eta^2 = .034$. Thus, the groups were comparable on the available physiological baseline measures, although causal conclusions remain limited by the nonrandomized design.

**Intervention Design**

The MECAI condition was developed in Unity using the Meta Interaction SDK, including passthrough and scene-understanding features that allowed the virtual interview setting to appear within the physical classroom. The embodied AI trainer was built with Convai, integrated into Unity through the Convai SDK and API key, and designed with Ready Player Me. The avatar included a seated interview role, gaze, gestures, and voice-based interaction; students used the Meta Quest Pro controller to initiate or interrupt speech through pressing the A button on the right controller. The ChatGPT voice condition used a structured prompt that instructed ChatGPT to act as an ESL job-interview trainer, ask one question at a time, allow learner errors, provide brief corrective feedback on grammar, vocabulary, and fluency, suggest improved professional responses, and adjust explanations to the learner's English level. The ChatGPT condition was designed using prompt engineering by the researcher. Both interventions AI personas, ChatGPT and Convai, were developed through prompt engineering by the researcher. The prompts provided the AI agent with specific contextual information, role definitions, behavioral expectations, and interaction constraints to ensure that responses remained aligned with the objectives of the job interview tasks. This approach enabled the researcher to guide the conversational flow and maintain consistency across learner interactions.

The researcher onboarded both experimental groups to the specific technologies used in their conditions. Participants in the ChatGPT group learned about audio-mode interaction and basic prompt literacy, while the MECAI group trained with Meta Quest Pro headsets and practiced engaging with the embodied AI agent. This initial training ensured everyone could work independently and prevented technical unfamiliarity from skewing our results.

**Job Interview Rubric and Inter-rater Reliability**

Interview performance was evaluated using a standardized analytic rubric developed for this study. The rubric assessed seven dimensions of performance: Fluency, Pronunciation, Intonation and Stress, Grammar and Sentence Structure, Vocabulary Use, Content and Relevance of Answers, and Use of Work-Based Examples. Scores on each dimension were summed to produce an overall interview performance score, with higher scores indicating better performance.

Three experienced ESL researchers independently evaluated participants' interview performances using the rubric. Inter-rater reliability was assessed using Fleiss' Kappa. According to the guidelines of Landis and Koch (1977), the obtained coefficient indicated substantial agreement among raters ($\kappa = .74$). Following reliability assessment, ratings were aggregated to produce a single overall performance score for each participant. The primary analysis focused on the overall interview performance score rather than individual rubric dimensions.

# 4.1 Statistical Analysis

A total of 75 students participated in the study, enrolled in Skills for Success level 2 course at Hudson County Community College, with mean (SD) age = 29.29 (12.96) years, with the majority being females (77.33%). The top participating nationalities were Dominican Republic (24.00%), Ecuador (18.67%), and

Egypt (14.67). The majority (85.33%) did not speak English at home. Over half (69.33) applied for a job in the US and 57.33% did not have formal training in English. Around (34.37%) felt neutrally confident when speaking English, 17.33% felt very confident, while only 5.33% felt very unconfident. The mean (SD) number of months of English learning and duration of residing in the US = 12.37 (19.13) and 38.57 (47.50), respectively. The mean (SD) number of spoken languages = 1.85 (0.71). The mean (SD) pre- and post-intervention total language performance scores = 8.75 (2.40) and 15.64 (3.21), respectively. The mean (SD) pre- and post-interview heart rates in beats per minute = 91.62 (17.80) and 88.66 (18.58), respectively. The mean (SD) pre- and post-intervention FLCAS = 89.28 (16.29) and 88.89 (17.94), respectively

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

Two-way repeated measures of ANOVA test was used to assess the difference between condition types (control, ChatGPT voice app, and MECAI) in the total language performance score, heart rate, and FLCAS across time and by the interaction between condition type and time, followed by Bonferroni adjustment if needed. Test statistics, degree of freedom (df), partial Eta squared ($\eta^2$), standard error (SE), estimated marginal means (EMM) and 95% confidence intervals (CI) were reported. *p* value < 0.05 was considered statistically significant.

# 5 Results

In repeated measures ANOVA, there was a significant difference between intervention types in the total language performance score ($F = 33.84$, $\eta^2 = 0.49$, $p < .001$) (Table 1). Total language performance scores were significantly highest in the MECAI group, followed by the ChatGPT and control groups (EMM [SE] = 13.94 [0.30], 12.24 [0.30], and 10.40 [0.30], respectively; $p < .001$) (Table 2, Appendix A, Table A1).

Difference between intervention types was non-significantly associated with heart rate ($F = 0.15$, $\eta^2 = 0.004$, $p = .86$) (Table 1), where the heart rate was non-significantly lowest in the MECAI group, followed by the control and ChatGPT groups (EMM [SE] = 89.20 [2.85], 89.76 [2.79], and 91.28 [2.79], respectively) (Table 2).

Difference between intervention types was non-significantly associated with FLCAS ($F = 1.75$, $\eta^2 = 0.05$, $p = .18$) (Table 1), where FLCAS was non-significantly lowest in the MECAI group, followed by the ChatGPT and control groups (EMM [SE] = 85.52 [2.62], 89.30 [2.62], and 92.44 [2.62], respectively) (Table 2).

Change in time was significantly associated with the total language performance score ($F = 298.20$, $\eta^2 = 0.81$, $p < .001$) (Table 1), where the post-test score was significantly higher than the pre-test score (EMM [SE] = 15.64 [0.26] and 8.75 [0.27], respectively) (Table 2).

Change in time was non-significantly associated with heart rate (F = 1.11, $\eta^2$ = 0.02, p = .30) (Table 1), where the post-test heart rate was non-significantly lower than the pre-test heart rate (EMM [SE] = 88.64 [2.17] and 91.53 [2.07], respectively) (Table 2). Likewise, change in time was non-significantly associated with FLCAS (F = 0.02, $\eta^2$ = 0.0003, p = .88) (Table 1), where the post-test FLCAS was non-significantly lower than the pre-test FLCAS (EMM [SE] = 88.89 [2.04] and 89.28 [1.89], respectively) (Table 2).

Post-test total language performance score was significantly higher than the pre-test score in all intervention types (F = 9.40, $\eta^2$ = 0.21, p < .001), yet the magnitude of change differed across groups. EMM (SE) of the pre- and post-test scores were 7.92 (0.47) and 12.88 (0.45), respectively, in the control group; 8.96 (0.47) and 15.52 (0.45), respectively, in the ChatGPT group; and 9.36 (0.47) and 18.52 (0.45), respectively, in the MECAI group (Tables 1 and 2; Appendix B, Fig. 2).

There was a non-significant difference in how heart rate changed across time between intervention types (F = 2.02, $\eta^2$ = 0.05, p = .14) (Table 1), where the EMM (SE) of the post-test heart rate was non-significantly higher than the pre-test heart rate in the control group (92.04 [3.73] and 87.49 [3.57], respectively); while the EMM (SE) post-test heart rate was lower than the pre-test heart rate in both the ChatGPT group (87.05 [3.73] and 95.51 [3.57], respectively) and the MECAI group (86.81 [3.81] and 91.58 [3.64], respectively) (Table 2, Appendix B, Fig. 3).

There was a non-significant difference in how FLCAS changed across time between intervention types (F = 0.85, $\eta^2$ = 0.02, p = .43) (Table 1), where the EMM (SE) of the post-test FLCAS was non-significantly higher than the pre-test FLCAS in the control group (92.60 [3.54] and 92.28 [3.28], respectively) and the ChatGPT group (90.92 [3.54] and 87.68 [3.28], respectively); while the EMM (SE) of the post-test FLCAS was lower than the pre-test FLCAS in the MECAI group (83.16 [3.54] and 87.88 [3.28], respectively) (Table 2, Appendix B, Fig. 4).

Table 1
Association between the intervention type, time and total language performance scale, heart rate, and FLCAS among participants (N = 75)

| Factor | Total language performance score | | | Hear rate | | | FLCAS | | |
|---|---|---|---|---|---|---|---|---|---|
| | F | $\eta^2$ | *p* value | F | $\eta^2$ | *p* value | F | $\eta^2$ | *p* value |
| **Group** | 33.84 | 0.49 | < 0.001* | 0.15 | 0.004 | 0.86 | 1.75 | 0.05 | 0.18 |
| **Time** | 298.20 | 0.81 | < 0.001* | 1.11 | 0.02 | 0.30 | 0.02 | 0.0003 | 0.88 |
| **Time*Group** | 9.40 | 0.21 | < 0.001* | 2.02 | 0.05 | 0.14 | 0.85 | 0.02 | 0.43 |

$\eta^2$ Partial Eta squared, FLCAS Foreign Language Classroom Anxiety Scale

Two-way repeated measures of ANOVA

*Statistically significant at *p* < 0.05

Table 2
EMM of the total language performance score, heart rate, and FLCAS in intervention types across time among participants (N = 75)

| Factor | | | Total language performance score | | Heart rate | | FLCAS | |
|---|---|---|---|---|---|---|---|---|
| | | | EMM (SE) | 95%CI | EMM (SE) | 95%CI | EMM (SE) | 95%CI |
| **Group** | **Control** | | 10.40 (0.30) [a] | (9.79, 11.01) | 89.76 (2.79) | (84.21, 95.32) | 92.44 (2.62) | (87.22, 97.66) |
| | **ChatGPT** | | 12.24 (0.30) [b] | (11.63, 12.85) | 91.28 (2.79) | (85.72, 96.84) | 89.30 (2.62) | (84.08, 94.52) |
| | **Multimodal AI** | | 13.94 (0.30) [c] | (13.33, 14.55) | 89.20 (2.85) | (83.52, 94.87) | 85.52 (2.62) | (80.30, 90.74) |
| **Time** | **Pre-test** | | 8.75 (0.27) | (8.21, 9.29) | 91.53 (2.07) | (87.39, 95.66) | 89.28 (1.89) | (85.51, 93.05) |
| | **Post-test** | | 15.64 (0.26) | (15.12, 16.16) | 88.64 (2.17) | (84.31, 92.96) | 88.89 (2.04) | (84.82, 92.97) |
| **Group*Time** | **Control** | **Pre-test** | 7.92 (0.47) | (6.98, 8.86) | 87.49 (3.57) | (80.37, 94.60) | 92.28 (3.28) | (85.75, 98.81) |
| | | **Post-test** | 12.88 (0.45) | (11.98, 13.78) | 92.04 (3.73) | (84.60, 99.49) | 92.60 (3.54) | (85.54, 99.66) |
| | **ChatGPT** | **Pre-test** | 8.96 (0.47) | (8.02, 9.90) | 95.51 (3.57) | (88.40, 102.63) | 87.68 (3.28) | (81.15, 94.21) |
| | | **Post-test** | 15.52 (0.45) | (14.62, 16.42) | 87.05 (3.73) | (79.60, 94.50) | 90.92 (3.54) | (83.86, 97.98) |
| | **Multimodal AI** | **Pre-test** | 9.36 (0.47) | (8.42, 10.30) | 91.58 (3.64) | (84.32, 98.84) | 87.88 (3.28) | (81.35, 94.41) |
| | | **Post-test** | 18.52 (0.45) | (17.62, 19.42) | 86.81 (3.81) | (79.21, 94.41) | 83.16 (3.54) | (76.10, 90.22) |

EMM Estimated marginal means, SE Standard Error, CI Confidence interval, FLCAS Foreign Language Classroom Anxiety Scale

Two-way repeated measures of ANOVA

[a–c] Different letters mean statistically significant difference between groups in the total language performance score

*Statistically significant at $p < 0.05$

# 5.1 Qualitative Findings: Student Perceptions Across Three Learning Conditions

Thematic analysis of structured interviews with students from three conditions — ChatGPT voice application, Mixed Reality Conversational AI Agent (MECAI in MR), and teacher-led instruction (control), indicated four overarching themes reflecting students' perceptions of their learning experiences.

# 5.2.1.Theme 1: How AI Shaped Confidence and Oral Communication

Students across both AI conditions reported generally positive first impressions, though initial reactions ranged from excitement to nervousness. Those in the MECAI in MR group described the technology as novel and engaging: '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.' Students in the ChatGPT group described early uncertainty that gave way to confidence: 'I felt a little nervous at first. I didn't know how it worked. But it was easy and interesting.' Teacher guidance during onboarding was a consistent factor in easing these early anxieties across both treatment groups.

For the ChatGPT group, three central subthemes emerged: personalization, a judgment-free practice environment, and the absence of physical embodiment. Students valued how ChatGPT tailored feedback to their needs and helped them prepare for real-world job interviews with relevant vocabulary and realistic questions. '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.' The AI's affirming and corrective tone was also appreciated: 'It was impact me that the AI Congrats me and correct me in a soft way.' However, students identified the lack of nonverbal cues as a meaningful limitation: '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.'

For the MECAI in MR group, the AI agent was frequently described in humanlike terms: a 'real person,' 'older friend,' or trusted advisor. Students noted that the embodied AI reduced anxiety compared to human interlocutors: '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.' Some students reported challenges with speech recognition, particularly due to accents and fast response speeds, but most adapted over time.

# 5.2 Theme 2: Immersion, Authenticity, and Sense of Presence in the MR Environment

A sub-theme specific to the MECAI in MR group focused on how the simulated office environment created a compelling sense of situational realism. Students described feeling present in a real interview setting, with the AI agent's eye-tracking and responsiveness contributing to this impression: 'That experience is different because I feel like I am talking with somebody... I can see that person.' Students

who had prior experience with ChatGPT found the embodied agent notably more realistic by comparison. One student highlighted a preference for the MR condition specifically because of this authentic quality: 'Yes. Very good. Feel real. I can see, talk. For me is best.' A small number of students mentioned environmental distractors within the MR scene, though this was not a dominant concern.

## 5.2.2. Theme 3: The Indispensable Role of the Human Instructor

Despite appreciation for AI tools, all three groups consistently affirmed the centrality of the human teacher. Students in the treatment groups credited the instructor with effectively integrating AI into the curriculum, making it accessible and purposeful. They distinguished between what AI could offer — personalization, repetition, specialized knowledge — and what only a human teacher could provide: emotional attunement, cultural guidance, and real-time responsiveness to nonverbal cues.

'One moment that helped me a lot was when we did a mock interview in class. You asked me a hard question... you helped me think of a personal example, and little by little I was able to answer.'

Students also acknowledged teacher limitations honestly: 'I think my teacher can give me ideas but not like ChatGPT. She doesn't have the same knowledge.' Some control group students noted that human instruction could occasionally increase anxiety due to fear of judgment or the teacher's mood. Nonetheless, all groups positioned AI as a supplement to, not a replacement for, teacher-led instruction: 'AI cannot do the work of a human. It needs to be used as a tool.'

## 5.2.3 Theme 4: Managing and Reframing Language Anxiety

Anxiety emerged across all groups as a nuanced experience. Students who practiced with AI consistently described lower fear of judgment compared to human interaction, with one student noting simply, 'Because it's AI, it's not human.' Both AI groups highlighted how repetition and personalization helped reduce anxiety over time, while students in the control group reflected that human instruction, though occasionally anxiety-inducing, also built confidence through encouragement and community. Interestingly, some students, across all groups, acknowledged that moderate anxiety could be productive: 'Anxiety can be healthy and motivating, guiding improvement when tailored instruction is provided.'

Students in the control group expressed interest in AI tools, with some saying they wished AI practice had been available to them. A student in the MR group articulated the broader finding well: 'Human good but I feel shy. ChatGPT has no face. I like Mixed Reality.' All groups converged on the view that a well-structured, teacher-guided environment in which AI plays a complementary role produced the most supportive learning conditions for reducing anxiety and building oral communication confidence.

# 6 Discussion

## 6.1 Research Question 1 and 2 Students' Speaking Performance and Anxiety

The findings of this study align with prior research showing that AI-supported conversational practice can improve second-language speaking performance, particularly when learners receive repeated, low-stakes opportunities to speak with responsive AI agents. The significant performance advantage for the MECAI mixed-reality group supports work on embodied conversational agents and XR language learning, which suggests that visible social presence, contextual cues, and simulated real-world interaction can strengthen oral performance more than decontextualized practice alone (Han & Ryu, 2025; Bendarkawi et al., 2025; Chun et al., 2025). The ChatGPT voice group also improved more than the control group, aligning with studies on AI conversation bots that show benefits for fluency, confidence, and speaking practice (Ding & Yusof, 2025; Lan & Zamri, 2025).

However, the anxiety results only partially align with prior research. Earlier studies often report reduced speaking anxiety after chatbot, VR, or AR-supported speaking practice (Chen, 2024; Kang, 2023; Ding & Yusof, 2025). In this study, FLCAS and heart-rate trends favored MECAI descriptively, but neither reached statistical significance. This is surprising given the strong performance gains and suggests that improved speaking ability may emerge before measurable reductions in anxiety. It may also reflect the high-stakes ESP/job-interview context, where learners can perform better while still feeling evaluated or physiologically aroused.

The strongest finding is that mixed reality with embodied conversational AI outperformed audio-only AI practice. This aligns with broader MR/AR research outside language learning, where embodied agents and spatially situated interaction often increase engagement, realism, and behavioral transfer compared with non-embodied or audio-only systems (Yang et al., 2025; Grubert & Sedelmaier, 2025). Compared with VR, MR may offer a useful middle ground: it preserves real-world environmental grounding while adding interactive virtual agents, potentially reducing the disorientation or artificiality sometimes associated with fully immersive VR. Thus, the MECAI advantage may come not only from AI conversation, but from embodied presence, visual feedback, and situated rehearsal.

Finally, the comparison between embodied conversational AI and audio AI apps suggests that embodiment matters. Audio-only tools such as ChatGPT voice can provide accessible, adaptive speaking practice, but embodied agents may better approximate interpersonal communication by adding gaze, gesture, turn-taking cues, and social presence. The mixed findings for anxiety indicate that embodiment may improve performance more reliably than affective outcomes, at least over six weeks. Future studies should examine whether longer exposure, larger samples, or more sensitive anxiety measures reveal stronger emotional benefits.

## 6.2 Research Question 3: Student Perceptions and Emerging Themes

Student perceptions across all conditions were largely positive, though each group valued different aspects of their experience. These qualitative findings reinforce a broadly emerging theme in the literature: that AI is most effective when integrated purposefully by a skilled human instructor rather than used as a standalone tool. Students' consistent elevation of teacher guidance, even when they acknowledged AI's superior content breadth, supports Warschauer and Healey's (1998) longstanding argument that technology in language learning serves its fullest function as a pedagogical supplement, not a replacement. The finding that control group students wished they had access to AI practice represents an emerging theme not widely addressed in prior literature and suggests that students themselves are increasingly aware of AI's supplementary value, independent of whether they were assigned to use it.

The anthropomorphization of the MECAI agent, described as a 'person,' 'woman,' 'older friend,' or 'someone', is a notable qualitative finding. This mirrors research on social robots and conversational agents that documents the human tendency to attribute social qualities to interactive agents (Nass & Moon, 2000), and suggests that embodied AI may carry affective and motivational benefits beyond its instructional content. This represents a meaningful area for future inquiry, particularly as generative AI agents become more sophisticated and their integration into language classrooms become more prevalent.

# 6.3 Limitation

Although social presence was not measured with a validated scale, the study did collect qualitative interview data about students’ experiences in the MR condition. These interviews provided insight into how students perceived the embodied AI agent, the immersive setting, and the realism of the job interview practice. However, because these data were self-reported and exploratory, they cannot fully rule out novelty, presence, or time-on-task confounds. Future studies should combine qualitative interviews with validated social presence measures and individual time-on-task logs.

The findings are consistent with social presence theory, which suggests that embodied conversational agents may create a stronger sense of interpersonal interaction than voice-only systems. However, social presence was not directly measured in the present study, and therefore this interpretation should be considered tentative. Future research should include validated measures of social presence to test this mechanism directly.

# 6.4 Practical Implications

The findings of this study have important implications for English for Specific Purposes (ESP) instruction, particularly in the area of job interview preparation for English language learners. The performance outcomes observed in the mixed reality (MR) embodied conversational AI condition suggest that immersive and interactive technologies that are pedagogically designed by human instructors can provide learners with authentic opportunities to practice workplace communication skills in a low-risk environment. Unlike traditional classroom activities, embodied conversational agents can

have the potential to simulate realistic interview interactions while providing learners with repeated opportunities for practice and feedback.

For educators, the results indicate that MR-based conversational AI may serve as a valuable supplement to conventional instruction by increasing opportunities for individualized speaking practice. Such technologies may be particularly useful in adult ESL and workforce development programs where learners often have limited opportunities to engage in authentic interview simulations, and instructors have limited knowledge in the professional area. The use of embodied agents may also help bridge the gap between classroom learning and real-world employment contexts by exposing learners to contexts that mimic realistic communicative demands.

For institutions and program administrators, the findings suggest that Pedagogically grounded and instructor-facilitated AI-supported interview training tools have the potential to improve learners' communicative readiness for employment while reducing reliance on extensive one-on-one instructor-led practice. As AI technologies become increasingly accessible, integrating conversational agents into ESP curricula may represent a scalable approach to supporting workforce-oriented language development.

As participants reported positive experiences with the AI-supported learning environments which encouraged them to take risks during language practice. This suggest that conversational AI may contribute to increased learner confidence during speaking practice, which are important factors in second language development and job interview preparation.

## 6.5 Future Research

Several directions for future research emerge from the present study. First, future studies can directly measure social presence, perceived realism, and learner engagement to better understand the mechanisms through which embodied conversational agents influence language learning outcomes. Although the present findings are consistent with social presence theory, these constructs were not directly assessed and therefore require empirical examination.

Second, future research should employ randomized controlled designs and larger, more diverse samples to strengthen causal inference and enhance the generalizability of findings. Including participants from multiple proficiency levels, educational settings, and cultural backgrounds would help determine whether the observed benefits extend across different learner populations.

Third, future studies should investigate the specific features of conversational AI systems that contribute most strongly to learning outcomes. For example, researchers may compare embodied and non-embodied agents while controlling for time-on-task, feedback quality, and interaction frequency. Such studies would help isolate the effects of embodiment, immersion, and conversational responsiveness.

Fourth, researchers should examine long-term learning outcomes and transfer effects. While the present study focused on short-term interview performance, future investigations could explore whether gains

are maintained over time and whether improvements transfer to authentic employment interviews and workplace communication contexts.

Finally, future work should incorporate additional objective and subjective measures of anxiety, including physiological indicators, self-report measures, and behavioral observations. A more comprehensive assessment approach may provide deeper insight into how AI-mediated learning environments influence both communication performance and affective outcomes among language learners.

## 7 Conclusion

This study examined the effects of traditional instruction, ChatGPT voice interaction, and Mixed Reality Embodied Conversational AI (MECAI) on ESL learners' job interview performance and speaking anxiety. The findings suggest that all three approaches supported improvements in oral performance, with the AI-supported conditions showing stronger observed gains than traditional instruction. Among them, the MECAI group showed the most notable improvement, indicating that embodied and immersive practice may offer useful support for ESP speaking development.

For speaking anxiety, the results showed a general trend toward reduction, particularly in the MECAI group; however, these changes were not statistically significant. Therefore, the findings should be interpreted cautiously and do not provide strong evidence that AI-mediated practice reduces anxiety within the six-week intervention period.

Overall, the study suggests that conversational AI, especially when embodied in mixed reality, may be a promising supplementary tool for job interview preparation. However, AI should be understood as a support to, rather than a replacement for, human instruction.

## Declarations

# Author Contribution

There's only one author of this manuscript who was solely responsible for obtaining IRB approval, obtaining letter of support from the community college where she worked at. She was also responsible for providing her won students with the Mixed Reality headset and onboarding to introduce the students to the technology and make them familiar with it. The researcher was also responsible for writing the manuscript, including all the sections and data analysis.

## References

1. Aliakbari, M., Barzan, P., Allahveysi, S. P., & Bakhtiarvand, M. (2025). AI-powered chatbots for English conversation practice: impact on speaking proficiency. *Technol Assist Lang Educ*, *3*(3), 45–62. https://doi.org/10.22126/tale.2025.12621.1138

2. Alwi, N. A. N. M., Shamsuddin, W. N. F. W., & Masrom, U. K. (2025). *Employing virtual reality-enhanced task-based language learning and teaching to mitigate interview anxiety*. Computer-Assisted Language Learning Electronic Journal.
3. Bagasala, G. L., & Estremera, M. L. (2025). A systematic literature review of task-based language teaching (TBLT) in L2 context. *International Journal for Multidisciplinary Research*, *7*(3), 1–12.
4. Bendarkawi, J., Ponce, A., Mata, S. C., Aliu, A., & Liu, Y. (2025). *Conversar: exploring embodied LLM-powered group conversations in augmented reality for second language learners*. ACM.
5. Cahya, N., & Churnia, E. (2026). Communication Anxiety and Speaking Performance in EFL Learners: A Systematic Review of Counseling and Pedagogical Strategies. *Indonesian Journal of Educational Counseling*, *10*(1), 74–86.
6. 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.
7. Chen, Y. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Comput Assist Lang Learn*. https://doi.org/10.1080/09588221.2024.000000
8. Chun, D., Kaplan-Rakowski, R., & Meyr, J. (2025). AI-mediated high-immersion virtual reality for language learning. *Language Education in Immersive Technologies*. Springer.
9. 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. *Humanit Soc Sci Commun*. https://doi.org/10.1057/s41599-025-00000-0
10. Du, J., & Daniel, B. K. (2024). Transforming language education: a systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Comput Educ Artif Intell*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230
11. Dülger, O., & Ugurlu, Z. Ö. (2025). *College graduates' foreign language needs in employment*. Journal of Pedagogical Research.
12. Fareen, J. A. M. (2017). *English for placement purposes: Developing a needs based course for meeting the employment demands of IT industries*. Eurasian Journal of Applied Linguistics.
13. Fernández, A. M. (2021). The effects of Task-Based Language Teaching on the Speaking Skill: A systematic research synthesis and meta-analysis. *Ciencia Digital*, *5*(4), 72–93.
14. Ghafar, Z. N., & Raheem, B. R. (2023). Factors affecting speaking proficiency in English language learning: a general overview on the speaking skill. *J Soc Sci*, *2*(6), 507–518.
15. Grubert, J., Sedelmaier, Y., & Landes, D. (2025, October). Towards embodied conversational agents for reducing oral exam anxiety in extended reality. In 2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 561–564). IEEE.
16. Guo, X., Jin, H., & Qi, T. (2023). How does social presence influence public crisis information sharing intention? Situational pressure perspective. *Frontiers in Public Health*, *11*, 1124876.

17. Han, D., & Ryu, J. (2025). *Exploring embodied conversational agents in extended reality for language learning*. Immersive Learning Research Network.
18. Huang, W., Hew, K. F., & Fryer, L. K. (2021). Chatbots for language learning—are they really useful? A systematic review of chatbot-supported language learning. *Journal Of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610
19. Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Routledge.
20. Kayhan, S. (2025). The Effect of Intervention Studies on Foreign Language Speaking Anxiety: A Meta-Analysis Study. *Journal of Pedagogical Research*, *9*(3), 62–83.
21. Kim, J., Merrill Jr, K., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383.
22. Musa, M. K., Rahim, A. A. A., & Khamis, N. Y. (2025). Enhancing Speaking Proficiency through Task-Based Activities: A systematic literature review in the TVET context. Environment-Behaviour Proceedings Journal, 10(SI33), 25–31.
23. Nghia, V. N. D., & Quang, N. N. (2021). Task-based language teaching to enhance learner communicative competence in Vietnam: a matter of opinion? *International Journal of Linguistics Literature and Translation*, *4*(8), 151–159.
24. Takaya, K. (2024). *Taiwanese EFL learners' use of learning strategies in developing job interview skills*. Journal of International and Comparative Education.
25. Tania, M., Danim, S., & Yensy, N. A. (2025). Foreign Language Speaking Anxiety: A Systematic Literature Review of Research Developments and Practical Implications. TPM–Testing, Psychometrics, Methodology in Applied Psychology, 32(3), 850–861.
26. Toncelli, R., & Kostka, I. (2024). A love-hate relationship: Exploring faculty attitudes towards GenAI and its integration into teaching. *International Journal of TESOL Studies*, *6*(3), 77–94.
27. Yu, Y., Mofreh, S. M., & Salem, S. (2024). Speak beyond borders: A systematic review of task-based language teaching for EFL speaking proficiency. *English language teaching*, *17*(7), 1–15.
28. Xiong, Y., Zhang, Q., Zhao, L., Liu, S., Guan, H., Sui, Y., Lee, K. M. R., et al. (2024). A meta-analysis and systematic review of foreign language anxiety interventions among students. *Journal of Language and Social Psychology*, *43*(5–6), 620–650.
29. Xu, T., Zhang, P., & Yin, J. (2026). Teaching Strategies for English Audio-Visual Oral Instruction Based on Deep Neural Networks. *International Journal of High Speed Electronics and Systems*, *35*(03), 2540456.
30. Zeng, S., & Della, C. (2024). *A systematic review of professional communication skills in English for Specific Purposes in vocational education within the new media era*. Eurasian Journal of Educational Research.
31. Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, *103*(4), 763–781.

# Figures

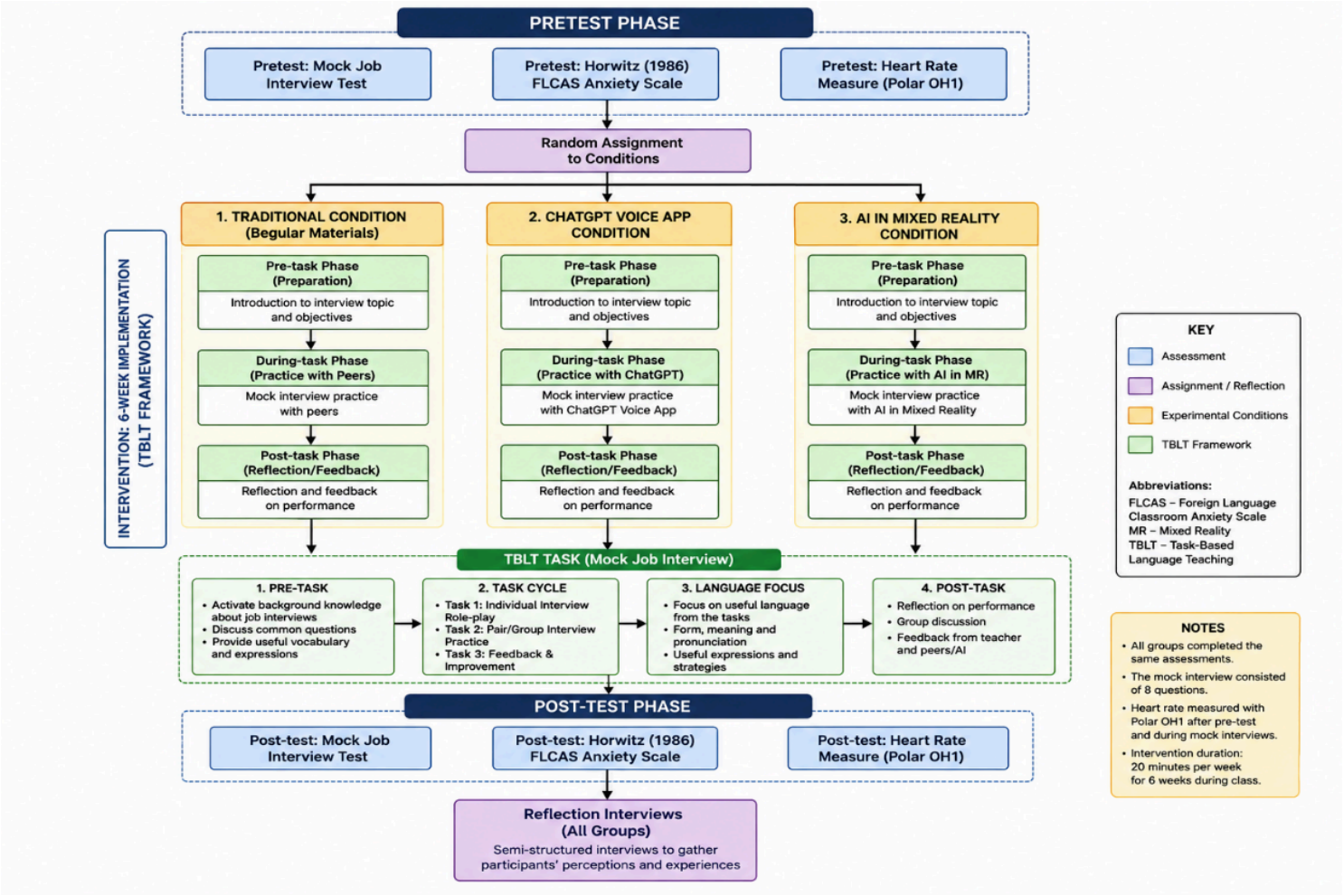



**Figure 1**

Legend not included with this version.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx