The JALT CALL Journal
ISSN 1832-4215, 22, 103142 (2026)
https://doi.org/10.29140/jaltcall.2026.103142

Castledown

An exploratory study of AI-avatar speaking practice in low-immersion VR: Fluency, anxiety, and technology acceptance

Yukie Saito
Chuo University, Japan
syukie001j@g.chuo-u.ac.jp

This exploratory mixed-methods study investigated whether speaking practice with AI-controlled avatars in low-immersion virtual reality (LiVR) on personal computers is associated with improved speaking fluency and reduced foreign language anxiety (FLA) among Japanese university EFL learners, and how learners perceive the platform. Eight students practiced speaking with AI-controlled avatars in Immerse at home for one month. Data included pre/post FLA surveys, audio recordings, reflective journals, and a post Technology Acceptance Model (TAM) survey. Results showed a small but statistically significant increase in fluency (words per minute) and a modest, statistically significant decrease in FLA. Qualitative data and TAM responses suggested LiVR practice was engaging, convenient, and emotionally supportive, while highlighting limits in conversational flexibility and occasional recognition problems. The fluency–FLA association was moderate but not statistically significant. Given the very small sample (N = 8), short duration, lack of a control group, and a single fluency metric, findings should be interpreted cautiously. Future controlled studies with larger samples and multidimensional fluency measures are needed to test causal effects and refine links among fluency, anxiety, and acceptance. As VR and generative AI evolve, AI-controlled avatars in LiVR may expand accessible interaction opportunities in EFL settings with limited authentic practice.

Keywords: AI-controlled avatars, low-immersion VR (LiVR), EFL speaking practice, speaking fluency (words per minute), foreign language anxiety (FLA), technology acceptance model (TAM)

Introduction

In today's globalized world, English is widely used for international communication, making speaking proficiency an essential skill for academic, professional, and social participation (Crystal, 2003; Kachru & Smith, 2008; Jenkins,

2009). However, many learners struggle to develop speaking ability due to limited opportunities for authentic interaction, speaking anxiety, and a lack of immersive environments (Horwitz et al., 1986; MacIntyre & Gardner, 1991; Tsiplakides & Keramida, 2009).

These challenges are particularly evident in English as a foreign language (EFL) contexts such as Japan (Yashima et al., 2004). A 2023 survey reported by MEXT (2024) indicates that nearly half of Japanese high school students did not reach CEFR A2 or higher. In many EFL classrooms, instruction prioritizes reading, writing, and grammar over speaking and listening (Nunan, 2003; Butler, 2011), and speaking tasks often lack authenticity because they are practiced in classrooms rather than real settings. Limited proficiency is also associated with higher foreign language anxiety, which can further inhibit oral performance (Thompson & Lee, 2012), whereas positive beliefs about language learning relate to lower anxiety and greater confidence (Aslan & Thompson, 2021). Supporting learners in reducing foreign language anxiety may therefore facilitate improvement in speaking.

One promising response is the use of virtual reality (VR), which can provide a safe, controlled space for speaking practice (Lin & Lan, 2015). In addition, advances in generative AI enable learners to interact with AI conversational partners, potentially increasing practice time and reducing pressure, especially when practice occurs at home. Although VR language learning and AI-based speaking practice have been studied, evidence on AI-controlled conversational avatars in low-immersion VR used at home remains limited, particularly in Japanese university EFL contexts. This study addresses this gap by examining whether one month of home-based low-immersion VR speaking practice with AI avatars is associated with changes in oral fluency (words per minute) and foreign language anxiety, and by exploring learner acceptance using the Technology Acceptance Model.

Theoretical frameworks

This study employs two complementary theoretical frameworks to investigate the effects of low-immersion VR interactions with AI avatars on speaking fluency and foreign language anxiety (FLA) among Japanese EFL learners: Foreign Language Anxiety (FLA) theory (Horwitz et al., 1986; MacIntyre, 1999) and the Technology Acceptance Model (TAM; Davis, 1989). FLA theory is adopted to examine language related anxiety associated with language use rather than technological stress, whereas TAM is used to examine students' perceptions of the AI avatar VR platform from a technological perspective rather than as a measure of language related anxiety.

According to Horwitz et al. (1986, p128), Foreign Language Anxiety (FLA) is defined as "a distinct complex construct of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process". MacIntyre (1999, p27) further define FLA as "the worry and negative emotional reaction aroused when learning or using a second language".

A primary cause of anxiety in speaking a foreign language is learners' fear or nervousness in communicating with others (He, 2018). Studies (Park & Lee, 2005; Hewitt & Stephenson, 2012) show that high levels of FLA negatively affect language learning outcomes, particularly speaking proficiency. In fact, existing literature about the effects of learning languages in VR suggests the potential of VR to lower FLA (Kaplan-Rakowski & Gruber, 2023; York et al., 2021; Park et al. 2025). However, Ding (2024) reported that despite positive perceptions of immersion, quantitative measures showed no statistically significant reductions in FLA, and Qiu et al. (2024) likewise found no overall significant effect, suggesting context-dependent outcomes. Moreover, most prior work has examined human-controlled avatars, and studies of VR-based speaking practice with AI-controlled avatars and its relationship to FLA remain limited.

In addition to FLA theory, this study incorporates the Technology Acceptance Model (TAM), introduced by Davis (1989), which is widely used to examine users' acceptance of and behavioral intentions toward new technologies. TAM proposes two primary constructs that influence users' attitudes toward and actual adoption of a technology: perceived usefulness and perceived ease of use. Perceived usefulness is the belief that a technology improves task performance, while perceived ease of use refers to how little effort is needed to use it. Extended versions of TAM have incorporated additional factors such as perceived convenience and perceived playfulness, offering a broader view of technology acceptance in educational settings (Chang et al., 2013). In the context of VR-based speaking practice, learners' perceptions of convenience, usefulness, and enjoyment are likely to influence their engagement with the tool. This study adopts TAM to assess students' views on the AI-avatar VR platform from a technology standpoint.

The rationale for adopting this combined theoretical framework is that FLA theory offers insights into the psychological and emotional factors, while TAM enables analysis of the technological factors that shape learners' acceptance and continued use of a VR-based speaking practice tool. To maintain conceptual clarity, this study distinguishes between the two frameworks: FLA addresses learners' psychological anxiety related to language use, whereas TAM focuses on their acceptance of and attitudes toward the technology. By integrating these two theoretical perspectives, this study investigates the relationship between speaking fluency, anxiety reduction, and learners' acceptance of practicing speaking with AI avatars in a VR platform.

Literature review

VR's definition, characteristics, and educational applications

Virtual reality (VR) is commonly defined as a computer-generated, immersive 3D environment that simulates realistic experiences, offering visual feedback that responds to users' movements (Dionisio et al., 2013). Through interactive simulations, VR enables participants to engage in environments closely resembling real-world scenarios, creating a strong sense of immersion. Due to

its immersive capabilities, VR is expected to have wide-ranging applications across domains such as shopping, entertainment, education, industrial training, healthcare, fitness, and social interaction (Hamad & Jia, 2022). Among these, numerous studies emphasize VR's significant potential in educational contexts (Hu-Au & Lee, 2017).

VR offers several educational benefits, including improved memory recall, increased intrinsic motivation, and enhanced empathy (Scrivner et al., 2019; Schwienhorst, 2002). It also promotes essential skills for the 21st century, such as critical thinking, creativity, empathy, and technology literacy (Hu-Au & Lee, 2017). Through immersive experiences, students can explore historical sites or scientific environments, making learning more engaging and accessible. Furthermore, VR enables safe, realistic training scenarios and promotes constructivist learning approaches, such as fostering empathy through simulations that allow users to adopt different perspectives (Bailenson et al., 2008). For example, Herrera et al. (2018) found that participants who experienced a VR simulation of becoming homeless developed more enduring compassionate attitudes and were more likely to support related social initiatives, compared to those who simply imagined being in other's situation without immersive technology.

VR-based language learning

VR has been extensively applied in educational contexts, and it has also shown promising effects in language learning. VR not only supports an integrated understanding of scientific concepts but also contributes positively to the retention of long-term memory. In particular, immersive virtual experiences hold significant potential for language education, where learner engagement and retention are key factors for success (Scrivner et al., 2019). The use of VR in language learning has also been linked to increased intrinsic motivation, enhanced intercultural awareness, and reduced affective barriers (Schwienhorst, 2002).

Research on technology-enhanced language learning has shown that virtual learning environments (VLEs) can support various aspects of language learning. Lin and Lan (2015) found that VLEs foster learner autonomy, confidence, creativity, and reduce anxiety by providing authentic, interactive contexts for practice – particularly valuable in foreign language settings where real-life opportunities are limited. Building on this foundation, recent studies suggest that virtual reality (VR) can further develop learners' speaking abilities. A meta-analysis by Qiu et al. (2024) reported a medium effect size ($g = 0.760$) for VR-assisted oral presentation activities, while Chen et al. (2022) observed the greatest improvements in speaking, followed by listening and affective factors. Supporting these findings, Ebadi and Ebadijalal (2022) demonstrated that VR-based collaborative tasks enhanced oral performance and willingness to communicate, and Liu and Hou (2020) found gains in fluency, pronunciation, and confidence through a mobile VR speaking application. However, Yudintseva (2023)'s review of 34 studies indicated that although VR use increased motivation and reduced anxiety, evidence for substantial gains

in speaking fluency remains limited. Taken together, these findings indicate a tension in the literature: VR tends to yield consistently positive affective outcomes (e.g., motivation and anxiety reduction), whereas evidence for measurable speaking gains remains mixed and context-dependent.

To better understand the varying effects of VR, it is important to consider the level of immersion provided by different types of VR technologies. VR is generally categorized into two types: high-immersion VR (HiVR) and low-immersion VR (LiVR). HiVR, accessed via head-mounted displays (HMDs), provides a fully immersive 3D environment, while LiVR uses standard 2D computer displays and offers lower immersion (Kaplan-Rakowski & Gruber, 2019). While prior studies highlight VR's educational benefits, HiVR tends to be more engaging and motivating than LiVR (Makransky & Lilleholt, 2018), but it also presents drawbacks such as motion sickness, eye fatigue, and higher equipment costs (Dolgunsöz et al., 2018; Kaplan-Rakowski & Wojdynski, 2018).

AI technology and AI avatars in VR

Artificial Intelligence (AI) is increasingly being used in English education to improve speaking skills. Recent studies show that AI-based instruction can enhance learners' oral proficiency, motivation, and self-regulation. For example, the study by Qiao and Zhao (2023) demonstrates that artificial intelligence-based instruction, using the Duolingo application, significantly enhances Chinese EFL students' speaking skills and self-regulation compared to traditional instruction. Similarly, Wei (2023) reported that AI-mediated English instruction using the Duolingo application significantly improved learners' language achievement, motivation, and self-regulated learning compared to traditional methods.

The rise of generative AI, especially ChatGPT, has further expanded opportunities for personalized and interactive language learning. Üstünbaş (2024)'s qualitative case study explored how Turkish pre-intermediate EFL learners in higher education used ChatGPT for speaking practice and demonstrated benefits such as enhanced speaking opportunities and scaffolding. Yıldız (2024) studied its effect on EFL learners' speaking self-efficacy, finding that students who practiced with ChatGPT significantly improved their confidence. Muniandy and Selvanathan (2024) conducted a study in a flipped classroom at a Malaysian university, finding that ChatGPT facilitated interactive speaking activities and contributed to improvements in learners' speaking proficiency. The study also reported that ESL learners perceived ChatGPT as user-friendly and effective in enhancing their speaking skills.

VR platforms like Mondly VR, Ovation VR, and Immerse VR allow learners to practice speaking with AI avatars in immersive environments. Although research remains limited, initial findings are promising. Kawasumi and Ishii (2023) compared VR and smartphone versions of Mondly with Japanese university students, finding that VR enhanced motivation but also caused greater fatigue. Godefridi et al. (2021) conducted an exploratory study indicating that combining Ovation VR and peer feedback may help reduce second language

learners' speaking anxiety and improve their confidence in public speaking situations. Thrasher et al. (2024) found that 64.1% of high school students who practiced speaking Spanish with AI avatars in Immerse VR enjoyed the experience, although some faced communication challenges due to the background noise from classmates sharing the same physical space. Despite generally favorable perceptions, prior studies also point to tensions such as limited interactional flexibility, fatigue, and environmental constraints (e.g., classroom noise), and many do not simultaneously report interpretable objective gains in speaking outcomes.

VR-based speaking practice can be particularly beneficial for EFL learners who lack opportunities for authentic communication. Previous studies have reported that learners perceive VR-based speaking practice with AI avatars as motivating, engaging, and enjoyable; however, relatively little research has examined whether such practice results in measurable improvements in speaking fluency or reductions in foreign language anxiety (FLA). In addition, the Technology Acceptance Model (TAM) has recently been applied in VR language learning research – for example, Barrett et al. (2023) highlighted immersion, enjoyment, curiosity, imagination, and interaction as key factors influencing learner acceptance. Nevertheless, TAM-informed investigations focusing specifically on AI-avatar–mediated VR speaking practice remain limited. Taken together, a key research gap is the scarcity of studies that simultaneously examine (a) measurable changes in speaking fluency, (b) FLA, and (c) learners' technology acceptance (e.g., via TAM) in low-immersion VR on personal computers (rather than HMD-based VR) with AI-controlled avatars, particularly in out-of-class or home contexts.

1. Does practicing speaking with AI avatars in VR improve students' speaking fluency?
2. Does practicing speaking with AI avatars in VR reduce students' foreign language anxiety?
3. How do students perceive the technological benefits and limitations of practicing English speaking with AI avatars in Immerse VR?
4. As an exploratory question, what possible relationships can be observed among changes in speaking proficiency, changes in foreign language anxiety, and students' perceptions of the technology?

Given the limited scope of this investigation, which involved a small sample (N = 8), the study adopts an exploratory approach. Accordingly, Research Questions 1, 2 and 3 are examined as primary inquiries, while Research Question 4 is treated as exploratory rather than confirmatory.

Methods

In order to address the research questions, this exploratory study employed a mixed-methods approach that integrated both quantitative and qualitative data. Quantitative data were collected from audio recordings of students' speaking practice with AI avatars, a Foreign Language Anxiety (FLA) questionnaire

administered before and after the practice, as well as from a post-questionnaire based on the Technology Acceptance Model (TAM). Qualitative data were collected from students' reflective journals, which provided insights into their experiences and perceptions of practicing speaking with AI avatars on a VR platform. This exploratory design was selected due to the small sample size (N = 8), and the study aimed to identify potential patterns rather than to produce generalizable findings. This section describes the VR platform used, the participants of the study, and the data collection and analysis procedures.

VR English lessons by Immerse Inc.

Immerse VR, developed by Immerse Inc., was the platform used for this study. It provides various immersive environments such as airports, restaurants, parks, and supermarkets, where learners can engage in English-speaking activities. In 2023, Immerse Inc. launched a new feature that allows students to practice speaking English with AI avatars through task-based interactions presented on the screen. Students can access Immerse VR via personal computers (PCs) and head-mounted displays (HMDs).

Although VR has been shown to support language learning, high-immersion VR (HiVR) can be more engaging and motivating than low-immersion VR (LiVR; Makransky & Lilleholt, 2018). However, HiVR also has drawbacks such as motion sickness, eye strain, and higher costs (Dolgunsöz et al., 2018; Kaplan-Rakowski & Wojdynski, 2018). A meta-analysis by Chen et al. (2022) reported that non-immersive VR had a larger effect (g = 1.121) on EFL learning than immersive VR (g = 0.347). Therefore, considering both the potential physical discomfort associated with HMDs, this study employed a low-immersion VR (LiVR) format, and students accessed Immerse VR via their PCs to practice speaking, without HMDs.

Additionally, in a study using Immerse VR, Thrasher et al. (2024) reported that students using the Meta Quest 2 device experienced communication difficulties with AI avatars due to the background noise caused by classmates in the same physical space. To address this issue, in the present study, each student participated individually from home to minimize external noise and ensure clearer interaction with AI avatars.

Participants and study design

Eight participants (aged 20–21, five males and three females) who were in an English communication course taught by the author at a private university in Tokyo participated in this study on a voluntary basis. Their English proficiency levels prior to this study were indicated by their TOEIC Speaking (S) and TOEIC Listening and Reading (L&R) scores, as shown in Table 1. Student A's TOEIC Speaking score could not be obtained. The TOEIC Speaking scores ranged from 60 to 140, with Students B and D having the highest scores (140) and Student G having the lowest score (60). For TOEIC Listening and Reading, the scores ranged from 405 to 910, with Student B achieving the highest score (910) and

Student F having the lowest score (405). Student B identified as a returnee with prior experience in an English-speaking country, whereas the others did not. None of the participants had prior experience using VR technology.

Baseline English proficiency was indexed by participants' most recent official Institutional Program (IP) TOEIC scores obtained through the university before the study. The variation in TOEIC scores reflects natural proficiency differences within the class rather than intentional selection. Due to this variation, analyses focused on within-subject pre–post changes instead of cross-participant comparisons, and potential ceiling effects were acknowledged as a limitation.

Table 1. Students' English levels before this study

Students' ID	Gender	TOEIC S score	TOEIC L&R score
A	M	–	555
B	M	140	910
C	M	80	410
D	M	140	715
E	M	110	555
F	F	90	405
G	F	60	670
H	F	90	525

Note. TOEIC Speaking and TOEIC Listening & Reading scores represent participants' most recent official IP test results obtained through the university before the start of the study. The researcher did not administer a separate TOEIC test. Partial data from two participants were previously reported as preliminary findings in *KOTESOL Proceedings 2024*.

The students engaged in English-speaking practice with AI avatars across nine different scenes provided in Immerse VR. In each scene, students interacted with three avatars, resulting in 27 speaking sessions per student (nine scenes × three avatars). This practice was conducted over a one-month period during winter break, when no formal English classes were offered. Participants were instructed to interact with one AI avatar per day during the one-month period, with each interaction lasting approximately three minutes, resulting in continuous daily speaking practice. Although individual pace and timing varied slightly, all participants completed the entire set of interactions by the end of the study.

Data collection and analysis

Audio-recorded data collection and analysis. The first and last speaking practice sessions with AI avatars were video- and audio-recorded via Zoom to assess improvements in speaking fluency. For consistency, the same VR scene (a bar) and the same task (greeting an AI avatar) were used in both practice sessions. The "bar" scene was selected as a basic setting for greetings, introductions, and casual exchanges, with no alcohol-related activities and a focus

solely on conversational practice for university students. Each speaking sample, which lasted approximately three minutes, provided short but comparable pre- and post-test utterances. The duration was standardized to ensure consistency in the range of spontaneous speech while maintaining participant engagement throughout the one-month period.

Although Immerse VR provides text-to-speech transcriptions of conversations, not all student utterances were transcribed correctly in the chat boxes. Therefore, the audio from the Zoom recordings of the first and last speaking practice sessions was manually transcribed by listening to it multiple times to accurately capture both the students' and the avatars' utterances. Speaking fluency was measured as Words Per Minute (WPM), calculated by dividing the total number of words spoken by the duration of speech (in minutes and seconds). Representative excerpts from pre- and post-practice transcripts of one participant, Student F, are presented in Appendix A. To enhance reliability, the author repeatedly listened to all recordings to verify transcription accuracy and ensure consistency in word counts and timing.

A paired-samples t-test was conducted to compare pre- and post-session WPM scores for fluency improvement. Due to the small sample size (N = 8) and potential non-normality, a Wilcoxon signed-rank test was also employed as an alternative method. This non-parametric test was selected to confirm the results without assuming a normal distribution of differences in WPM scores. Both statistical tests were two-tailed with $\alpha = .05$, and effect size (Cohen's *d*) was calculated to indicate the magnitude of change. These procedures were selected to ensure transparency and interpretability despite the limited statistical power associated with a small sample.

FLA questionnaire collection and analysis. The FLA questionnaire adapted from Foreign Language Classroom Anxiety Scale (FLCAS) (Horwitz et al., 1986) was administered before and after the speaking practice sessions. The original questionnaire by Horwitz et al. (1986) contains 33 items; however, nine items (Items 2, 5, 8, 11, 14, 18, 22, 28, and 32) that were considered more related to lower foreign language anxiety were excluded from this study, resulting in 24-items. The FLA questionnaire is provided in Appendix B. A five-point Likert scale was employed, with 5 for strongly agree, 4 for agree, 3 for neither agree nor disagree, 2 for disagree, and 1 for strongly disagree. Mean scores of pre- and post-questionnaire responses were compared using paired-samples t-tests to evaluate statistical changes in FLA scores. To confirm internal consistency, Cronbach's alpha was calculated for both administrations (pre = .91, post = .89), indicating high reliability.

Students' journals and perception questionnaire collection and analysis. During the sessions, the students were asked to keep reflective journals about their experiences and perceptions of the VR-based speaking practice. The journal included the scene, task details, and space for post-task comments. At the end of the final session, they also provided overall feedback on the experience, including positive and negative points, technical issues, and their

views on whether the practice was helpful for improving their speaking skills. Qualitative data from the journals were analysed using thematic analysis. The entries were thoroughly read multiple times for familiarization, and relevant segments were coded. The author conducted the coding and thematic analysis, repeatedly reviewing the journal entries to ensure accuracy and consistency in identifying themes. The themes were refined through iterative review to enhance confirmability and minimize subjective bias.

In addition, the students completed a questionnaire designed to assess their perceptions of practicing speaking with AI avatars in Immerse VR. The questionnaire was based on the Technology Acceptance Model (TAM) (Davis, 1989), including factors of perceived convenience, perceived playfulness, perceived ease of use, perceived usefulness, and continuance intention to use (Chang et al., 2013). Example items included "I can learn English at any time with Immerse," and "Immerse is convenient for me to engage in practicing speaking English." The questionnaire is provided in Appendix C. A five-point Likert scale was employed, with 5 for strongly agree, 4 for agree, 3 for neither agree nor disagree, 2 for disagree, and 1 for strongly disagree. The questionnaire was conducted after the final speaking practice with AI avatars.

Comparative analysis of WPM, FLA, and TAM scores. To explore the relationships among improvements in speaking fluency (measured by WPM), reductions in foreign language anxiety (FLA scores), and students' perceptions of the technology (TAM scores), correlation and regression analyses were conducted. Pearson correlation coefficients were calculated to assess the relationships between these variables. Furthermore, simple linear regression analyses were conducted to explore the extent to which changes in speaking fluency (WPM difference) could predict reductions in anxiety (FLA reduction) and students' perceptions of the VR platform (TAM scores). Given the very small sample size (N = 8), these analyses were interpreted cautiously, and no inferential claims were made for non-significant results. Instead, correlation magnitudes were reported descriptively to indicate possible directions for future research rather than to imply confirmatory evidence.

Results

To explore whether practicing speaking with AI avatars in VR improves students' speaking fluency and reduces foreign language anxiety, and how students perceive the technological benefits and limitations of practicing English speaking with AI avatars in Immerse VR, audio data, the FLA questionnaires, students' reflective journals, and the TAM questionnaire were analyzed. In addition, relationships among changes in speaking fluency, foreign language anxiety, and students' perceptions were examined. This section presents results from the audio data, the FLA questionnaires, the journals, and the students' perception questionnaire.

Results of video analysis

The effectiveness of speaking practice with AI avatars in LiVR was evaluated by comparing Words Per Minute (WPM) in the first (Pre) and last (Post) practice sessions. The mean Pre-WPM was 115.98 (SD = 32.88), whereas the mean Post-WPM increased to 139.26 (SD = 21.71). On average, students gained 23.29 WPM (SD of differences = 18.64). Notably, Student B, whose Pre-WPM was 183.5, showed no improvement (Gap = 0), while Student F achieved the largest gain (+58.9 WPM). Pre-time and Post-time are reported in minutes and seconds (mm:ss) to indicate the total speech time of each student at the initial and final sessions, respectively; detailed results for all eight students are presented in Table 2.

Table 2. Individual WPM scores and pre- and post- differences

Students' ID	Pre-time (mm:ss)	Pre-test word count	Pre-WPM	Post-time (mm:ss)	Post-test word count	Post-WPM	WPM gap
A	0:29	44	91.03	0:35	69	118.29	27.26
B	0:17	52	183.53	0:17	52	183.53	0.00
C	0:32	72	135.00	0:30	69	138.00	3.00
D	0:53	92	104.15	0:50	101	121.20	17.05
E	0:56	90	96.43	0:43	88	122.79	26.36
F	1:14	108	87.57	0:34	83	146.47	58.90
G	0:22	49	133.64	0:11	28	152.73	19.09
H	0:51	82	96.47	0:27	59	131.11	34.64

Note. WPM = words per minute. Mean Pre-test WPM = 115.98 (SD = 32.88); Mean Post-test, WPM = 139.26 (SD = 21.71); Mean WPM Gap = 23.29 (SD = 18.64).

To assess whether practicing speaking with AI avatars in VR led to a significant improvement in speaking fluency, a paired-samples *t*-test was conducted to compare Pre-WPM and Post-WPM scores. The analysis showed a statistically significant increase, $t(7) = 3.53$, $p = .010$, 95% CI [7.70, 38.87], with a large effect size (Cohen's $d = 1.25$), indicating that the practice in Immerse VR may have enhanced participants' speaking fluency.

Given the small sample size (N = 8) and potential deviations from normality, a Wilcoxon signed-rank test was also performed. The nonparametric test likewise revealed a significant difference in WPM, $W = 0.0$, $p = .014$ (Pratt's method). The agreement between the two analyses provides tentative support for the observed improvement, though the evidence remains limited.

Although the increase reached statistical significance, the numerical gain (approximately 23 WPM) represents only a modest practical improvement. Thus, while the data suggest a possible positive effect of the VR-based practice on speaking fluency, this finding should be regarded as preliminary, given the short intervention period and small sample size.

Results of FLA questionnaires

This study investigated changes in foreign language anxiety (FLA) among the eight students by comparing their responses to the FLA questionnaire administered before and after the practice sessions. The mean FLA score prior to the practice sessions (Pre-FLA) was 3.04 (SD = 0.73), while the mean score following the sessions (Post-FLA) was 2.78 (SD = 0.75), indicating an average reduction of 0.27 points. A paired-samples t-test showed that this decrease was statistically significant, $t(7) = 2.55$, $p = .038$(negative t indicates Post-FLA < Pre-FLA). The calculated effect size (Cohen's $d = 0.90$) indicates a large statistical effect. Although the difference reached statistical significance, the numerical change (0.27 points) was small, suggesting that the observed reduction in anxiety should be interpreted as limited in practical terms. The result therefore points to a measurable but modest improvement rather than a substantial decrease in FLA.

An item-level analysis was performed to determine the specific aspects of anxiety that were most influenced. The greatest reductions were observed in Item 1 ("I never feel quite sure of myself when I am speaking in my foreign language class") and Item 7 ("I keep thinking that the other students are better at languages than I am"), both showing a mean decrease of 0.75 points. A smaller decrease was found in Item 6 ("During language class, I find myself thinking about things that have nothing to do with the course"), which declined by 0.38 points. Conversely, a slight increase in anxiety was observed in Item 4 ("It frightens me when I don't understand what the teacher is saying in the foreign language"), where the mean score increased by 0.50 points. In addition, increases in standard deviations on several items indicate greater variability in student responses following the speaking practice with AI avatars. These mixed item-level results suggest that overall anxiety decreased, but not all aspects of anxiety were equally reduced.

To further explore individual differences, each participant's Pre- and Post-FLA means were examined. As shown in Table 3, Student F demonstrated the greatest reduction in anxiety, with a 0.67-point decrease. Substantial decreases were also observed for Student E (−0.54) and Student B (−0.46). In contrast, Student D experienced a slight increase in anxiety (+0.25), which may reflect normal individual variation rather than a systematic effect. Given the small sample, such differences should not be generalized but instead treated as exploratory observations.

Table 3. Results of individual students' FLA changes

Students' ID	Pre-FLA mean	Post-FLA mean	FLA change (post – pre)
A	2.88	2.71	−0.17
B	1.50	1.04	−0.46
C	3.58	3.46	−0.12
D	2.83	3.08	+0.25
E	3.25	2.71	−0.54
F	3.88	3.21	−0.67
G	2.92	2.83	−0.09
H	3.50	3.17	−0.33

Qualitative results of students' perceptions of Immerse VR

A thematic analysis of the students' reflective journals revealed generally positive perceptions of speaking practice with AI avatars in VR. Many participants described the experience as engaging, realistic, and helpful for building speaking confidence. A closer examination of the journals and final feedback identified three recurring themes regarding students' experiences with AI avatars and low-immersion VR: confidence building and reduced anxiety, authenticity and engagement, and technical and interactional challenges.

The first theme, confidence building and reduced anxiety, appeared across most students' reflections. Learners reported that interacting with AI avatars allowed them to speak more freely and comfortably in English without fear of judgment. Student F noted that speaking with avatars was easier than speaking with real people. Similarly, Student H appreciated being able to self-correct with textual feedback, describing the experience as a safe environment for practicing English.

The second theme, authenticity and engagement, reflected students' perceptions of the practice as realistic and motivating. Many participants emphasized that the virtual settings and tasks closely resembled real-world situations, which made the experience both practical and enjoyable. Student B described the interaction as "close enough to having a conversation with a real person," and Student E commented that it "felt like living in a foreign country." Student G also appreciated that the scenarios mirrored daily life.

The third theme, technical and interactional challenges, concerned issues with system usability and recognition accuracy. Some students mentioned that avatar movement was slow or that responses were occasionally delayed. For instance, Student E found it "difficult to move the character on the device," and Student D noted that more responsive controls would make the experience feel smoother. Several students also commented that the AI avatars lacked conversational flexibility; for example, Student B observed that the avatars only discussed assigned topics, which made interactions feel less natural and limited opportunities for spontaneous expression. In addition, some participants experienced issues with the system's voice recognition accuracy. Student A reported

that the AI occasionally had "difficulty hearing," while Student G noted that some of the student's pronunciations were not recognized. To maintain conceptual clarity, these technical and interactional issues were interpreted as aspects of system usability and technological acceptance rather than indicators of foreign language anxiety. Accordingly, challenges such as slow avatar movement and speech-recognition errors were treated as reflections of system performance rather than emotional reactions. A few students even suggested that adapting to such irregularities mirrored the unpredictability of real-life conversations, thereby enhancing realism.

Quantitative results of students' perceptions of Immerse VR

In addition to qualitative data, the post-session questionnaire measured students' perceptions of Immerse VR using a Technology Acceptance Model (TAM) framework. The questionnaire included 18 items assessing perceived usefulness, perceived ease of use, and behavioral intention to continue using the system, rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

The mean TAM scores of the individual students ranged from 3.56 (Student H) to 4.44 (Student F), indicating relatively consistent positive perceptions across participants (M = 3.90, SD = 0.28). Most students reported moderately high acceptance levels (Student A = 4.06, B = 3.72, C = 3.89, D = 3.89, E = 3.94, F = 4.44, G = 3.72, and H = 3.56). Notably, Student F, who achieved the largest improvement in speaking fluency (+58.9 WPM), also reported the highest TAM score. Conversely, Student H, despite describing the VR environment as a safe and supportive space in their reflection, recorded the lowest TAM score (3.56), perhaps reflecting frustration with technical usability rather than negative attitudes toward the learning experience itself.

Items related to convenience – such as perceiving Immerse as a practical and accessible tool for English practice – received the highest ratings (Q3 and Q4: M = 4.50, SD = 0.53), while flexibility-related items ("I can learn English at any place with Immerse") received the lowest (M = 3.00, SD = 0.93). These results suggest that participants evaluated Immerse VR as a convenient and engaging platform for speaking practice; however, the lower score for flexibility likely reflects the fact that the system was accessed via PCs rather than mobile devices, limiting its portability and use in different locations.

Overall, students evaluated Immerse VR positively, appreciating its convenience and engaging features while acknowledging occasional technical limitations. Individual differences in TAM scores (ranging from M = 3.56 to 4.44) appeared modest and likely reflected variations in digital familiarity rather than meaningful disparities in technological acceptance.

Comparative analysis of WPM, FLA and TAM scores

The relationships among the students' improvements in speaking fluency (WPM Difference), reductions in language anxiety (FLA Reduction), and their

perceptions of Immerse VR (TAM Scores) were analyzed using Pearson correlations and simple regressions (N = 8, two-tailed tests).

The relationship between WPM Difference and FLA Reduction showed a moderate negative correlation (r = –.467, p = .243), although the result was not statistically significant. The corresponding regression indicated that WPM Difference explained 21.8 % of the variance in FLA Reduction (R^2 = .218). Similarly, the correlation between TAM Scores and FLA Reduction was weakly negative (r = –.357, p = .385); the regression analysis showed that TAM Scores explained 12.8% of the variance in FLA Reduction (R^2 = .128), confirming a weak, non-significant relationship. In contrast, the correlation between WPM Difference and TAM Scores was moderate and positive (r = .612, p = .107), with the regression model accounting for 37.5 % of the variance (R^2 = .375), but this result was also not statistically significant. Given the very small sample size (N = 8), these results should be interpreted with caution. None of the relationships reached statistical significance, and therefore no reliable associations among fluency gains, anxiety reduction, or technology acceptance were confirmed. Nevertheless, the direction of the coefficients is consistent with theoretical expectations, suggesting possible patterns that could be examined more rigorously in future research with a larger sample.

Discussion

This study investigated the impact of AI avatar-based virtual reality (VR) speaking practice on the learners' speaking proficiency, foreign language anxiety (FLA), and their perceptions of the Immerse VR. As this exploratory study involved a small sample, its findings should be viewed as preliminary. Based on the results of both quantitative and qualitative data, the discussion addresses key findings: improvements in speaking fluency, changes in FLA, the learners' perceptions of the VR environment, and the potential interrelationships among these variables. Additionally, the findings are discussed through the lens of two theoretical frameworks: FLA and TAM. The final part of the discussion provides pedagogical implications and recommendations.

Improvements in speaking proficiency

The results suggest that VR-based speaking practice with AI avatars may contribute to improvements in speaking fluency, as indicated by an average increase of 23.29 Words Per Minute (WPM). The observed effect size (Cohen's d = 1.25) also indicates a potentially meaningful impact. These findings are consistent with the previous study (Ebadi & Ebadijalal, 2022) that has reported gains in speaking performance following VR-based interventions. Although the increase in speaking proficiency was statistically significant, the numerical gain was modest. Taken together, the results can be described as statistically significant but limited in practical terms, rather than as a large and meaningful effect. Thus, although these findings indicate the potential of practicing speaking with AI avatars in VR environments for improving speaking fluency,

additional research is necessary to determine their applicability across various contexts and learner populations.

In addition, although most students demonstrated measurable gains, individual variation was observed. For example, Student B, whose initial English proficiency was high and the pre-WPM score was relatively high, did not show further improvement. In his reflective journal, Student B wrote that the avatars primarily engaged with assigned topics, which made him feel that the conversations were less natural and less personal. This comment suggests that relatively highly proficient English speakers like Student B may need conversations extending beyond assigned topics. Such findings are in line with Yudintseva (2023), who reported that the effects of VR on fluency development can vary substantially among learners. Having AI avatars engage in open-ended conversations rather than assigned topics might be more beneficial for relatively high English- proficiency students like Student B.

Reduction in foreign language anxiety

The study also found a statistically significant decrease in FLA, with an average reduction of 0.27 points on the FLA questionnaire. This supports existing literature emphasizing the anxiety-reducing potential of immersive environments (Kaplan-Rakowski & Gruber, 2023; Thrasher, 2022). The students' journals showed a similar trend, with several participants indicating that they felt more at ease talking with AI avatars than with real people. Less pressured interactions with AI avatars in VR may have created an emotionally safer context for speaking, consistent with Lin and Lan's (2015) findings on reduced emotional filtering in VR settings.

However, the effect on anxiety varied across participants. One student (Student D) showed a slight increase in FLA. Item-level analysis indicated that specific types of anxiety, such as the fear of misunderstanding teacher input, were more prominent in some learners. This variability aligns with findings by Qiu et al. (2024), who observed no significant overall effect of VR on FLA, suggesting that factors such as learning context, system design, and individual differences may influence outcomes. While immersive environments have the potential to reduce foreign language anxiety, they can also introduce new challenges, such as the need to be familiar with the VR system, which may trigger anxiety. Some participants reported technical difficulties, including slow avatar responses, navigation challenges, and inaccurate voice recognition. These technical frustrations should be viewed as issues of system usability rather than as components of FLA per se, although they may have temporarily heightened learners' anxiety during interaction.

Student perceptions and technology acceptance

Qualitative and quantitative data revealed generally positive perceptions of Immerse VR. Students appreciated the opportunities of speaking practice with AI avatars, supporting the claim that authentic interactions enhance learner

engagement (Scrivner et al., 2019). These perceptions were reflected in the TAM questionnaire results, which showed high average ratings for convenience and engagement. Individual differences were minimal, as expected for a small sample. Consequently, the analysis emphasized overall positive trends rather than minor variations among participants.

At the same time, several technical limitations were commonly noted by participants, including voice recognition issues and the restricted nature of conversational flow. These challenges are in line with limitations reported in the previous research of speaking with AI avatars in VR (Kawasumi & Ishii, 2023; Thrasher et al., 2024). Such comments were interpreted as reflections of system usability within the TAM framework, not as indicators of foreign language anxiety. For VR platforms to be more effective in supporting speaking practice with AI avatars, improvements in AI conversational adaptability and system responsiveness are needed.

It is also important to consider how the implementation context can shape learners' experiences. Thrasher et al. (2024), for example, reported communication difficulties in the HiVR classroom setting, where the background noise from other students interfered with avatar interactions. In contrast, the present study used a low-immersion VR (LiVR) format, with the participants accessing the VR platform individually from home, thereby minimizing environmental distractions. This difference may have contributed to smoother interactions and more consistent user satisfaction.

Relationships among fluency gains, anxiety reduction, and perceptions

Although the correlation and regression analyses did not produce statistically significant results, certain patterns emerged. The moderate negative correlation between WPM improvement and FLA reduction suggests a potential link between increased fluency and reduced anxiety. However, because the correlations were non-significant and based on a very small sample (N = 8), these should not be described as moderate trends but rather as possible directions for future investigation. Although not conclusive, this tendency aligns with the theoretical assumption that higher communicative competence can reduce anxiety in language use (MacIntyre & Gardner, 1991).

Similarly, the positive correlation between WPM improvement and TAM scores indicates that learners who experienced greater gains also perceived the platform more favorably. This finding aligns with the Technology Acceptance Model (Davis, 1989), which indicates that perceived usefulness significantly influences user attitudes. Again, because the results were not statistically significant, they should be presented descriptively rather than inferentially. Although none of the relationships reached statistical significance, possibly due to the small sample size, the trends indicate a promising direction for future research with a larger group of participants.

Interpretations of theoretical frameworks

This study adopted two frameworks, FLA and TAM. FLA was used to examine whether speaking practice with AI–controlled avatars in low-immersion VR can reduce learners' foreign language anxiety. While prior VR language-learning research suggests that VR may lower FLA (Kaplan-Rakowski & Gruber, 2023; York et al., 2021; Park et al., 2025), quantitative findings remain mixed (Ding, 2024; Qiu et al., 2024), indicating context-dependent outcomes. Given the limited research on AI-controlled avatars, applying FLA provides a theory-based basis for evaluating anxiety change and for addressing the multidimensionality of the construct; consistent with this, the results showed an overall decrease in anxiety, alongside non-uniform item-level and individual shifts. However, overall, the findings suggest that speaking practice with AI-controlled avatars in LiVR warrants further examination as a potential approach in EFL contexts where opportunities for spoken interaction are limited and may be associated with persistently high FLA.

TAM was used to assess learners' technology-focused perceptions of the platform (e.g., usefulness, ease of use, convenience, enjoyment, and continuance intention) and to identify constraints that may shape engagement. However, TAM ratings alone may not sufficiently account for the nature and sources of specific implementation issues; more fine-grained problems (e.g., occasional recognition errors, inflexible exchanges, or unexpected system responses) are more effectively captured through reflective journals. In this study, the relationship between TAM and changes in FLA was also examined; however, it was weak, suggesting that favorable platform perceptions were not necessarily accompanied by greater anxiety reduction. One possible reason is that system-related issues (e.g., recognition errors or unexpected responses) may have elicited technology-related anxiety or frustration distinct from FLA. This possibility calls for further investigation of not only technology acceptance but also technology-related anxiety when learners engage in LiVR-based speaking practice with AI-controlled avatars.

Pedagogical implications and recommendations

The findings suggest two practical implications for supporting speaking development in EFL contexts. First, incorporating AI avatars in VR environments can provide learners with low-anxiety, self-paced speaking practice opportunities, which are particularly valuable in contexts where real-life interaction is limited in EFL contexts. Practicing speaking with AI avatars in VR may be particularly beneficial for lower-English proficiency learners by allowing them to lower language anxiety before engaging in authentic communication. Given this context, the integration of AI avatar-based speaking practice in VR may be particularly useful in Japan, where learners often struggle with speaking proficiency despite years of formal English education, as reflected in the 2023 MEXT survey showing that nearly half of the high school students did not reach a CEFR A2 level or higher.

Second, the results indicate that low-immersion VR (LiVR) on personal

computers can support meaningful language practice, suggesting that high-immersion VR (HiVR) on head-mounted displays (HMDs) is not a prerequisite for engaging AI-avatar speaking tasks. From a feasibility standpoint, LiVR is more accessible because it is web-based and runs on standard laptops, whereas HiVR typically requires HMDs (e.g., the Meta Quest 3S, costing around 50,000 JPY per unit) and adequate physical space. LiVR may also be better suited to repeated use, as it is often associated with less eye strain and fatigue than HiVR. Nevertheless, successful implementation depends on stable internet access, smooth management of multiple logins, and teacher support for designing and supervising AI-avatar speaking tasks. While home use may reduce noise and performance pressure, classroom use may also be feasible if logistics are planned – for example, by scheduling short individual speaking slots with AI avatars and integrating them into regular lesson routines.

Limitations and future directions

While this study offers valuable insights, several limitations should be acknowledged. First, the speaking samples used for the fluency analysis in the pre- and post-intervention recordings were brief, often lasting less than one minute, and the same task in the same scene was used for both assessments. Future studies should use longer samples and more diverse tasks across multiple scenes to reduce task familiarity effects and to capture a broader range of communicative performance. In addition, speaking fluency was operationalized using words per minute (WPM) only, which provides a limited view of fluency. Future research should therefore incorporate multidimensional indices (e.g., breakdown, speed, and repair fluency) and, where possible, standardized speaking assessments administered outside the VR platform (e.g., TOEIC Speaking, IELTS) to evaluate speaking ability more comprehensively. Second, the study employed a single-group design without a control group, which limits the ability to attribute observed changes solely to the VR intervention. Future studies should adopt randomized controlled or comparison-group designs to clarify causal relationships and rule out alternative explanations such as repeated testing effects. Third, the very small sample size (N = 8) constrains generalizability and limits statistical power, increasing the likelihood of unstable estimates and making it difficult to detect effects reliably. Although promising trends were observed, larger-scale studies are needed to more robustly examine relationships among fluency, anxiety, and technology-related perceptions. Finally, the short duration of the study limits conclusions about longer-term effects. Longitudinal research is needed to determine whether improvements in fluency and reductions in anxiety are sustained over time and whether learners' engagement with AI-avatar practice changes as novelty effects diminish.

Conclusion

This exploratory study examined the effects of AI avatar-based speaking practice in a low-immersion VR environment on EFL learners' speaking development,

FLA, and perceptions of the technology. The findings indicated measurable improvements in speaking fluency and a modest reduction in FLA. Learners also reported generally positive experiences, describing the VR environment as realistic, engaging, and emotionally supportive. However, consistent with the exploratory nature of this study, the results should be interpreted cautiously, with an emphasis on possible rather than confirmatory effects. While the correlations among speaking gains, anxiety reduction, and technology acceptance were not statistically significant, the directions of the relationships were generally consistent with theoretical expectations. These patterns should be viewed as tentative and require confirmation in future research.

Regarding changes in speaking proficiency, although this study assessed fluency solely in terms of changes in words per minute (WPM), not only fluency but also accuracy and complexity need to be investigated by tracking changes in accuracy as well as grammatical and lexical complexity over time. Building on these production-based analyses, subsequent work could then examine interactional processes in greater detail – such as turn-taking patterns, negotiation of meaning, and Initiation–Response–Follow-up (IRF) sequences – to clarify the extent to which learner–AI-avatar exchanges are associated with the development of communicative competence. In addition, although this study did not analyze the utterances produced by the AI avatars, future research could examine pragmatic features of AI avatars' utterances – such as the provision of feedback, the use of backchannels, and follow-up questioning – to better understand how these interactional moves might shape learners' willingness to communicate and speaking development. As VR and generative AI technologies continue to advance, their integration into language education may offer opportunities to create supportive, low-anxiety environments, particularly in EFL contexts where opportunities for authentic interaction are limited.

Declaration of use of generative AI

This manuscript made use of ChatGPT to assist in paraphrasing and improving the clarity of text and ChatGPT for the analysis of quantitative data in Tables 2 and 3. No AI tools were used for the generation of research hypotheses or the writing of conclusions. The final manuscript was authored and reviewed by the listed author without additional AI input.

References

Aslan, E., & Thompson, A. S. (2021). The interplay between learner beliefs and foreign language anxiety: Insights from the Turkish EFL context. *The Language Learning Journal, 49*(2), 189–202. https://doi.org/10.1080/09571736.2018.1540649

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences, 17*(1), 102–141. https://doi.org/10.1080/10508400701793141

Barrett, A., Pack, A., Guo, Y., & Wang, N. (Joanne). (2023). Technology acceptance model and multi-user virtual reality learning environments for Chinese language education. *Interactive Learning Environments, 31*(3), 1665–1682. https://doi.org/10.1080/10494820.2020.1855209

Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics, 31*, 36–57. https://doi.org/10.1017/S0267190511000122

Chang, C.-C., Liang, C., Yan, C.-F., & Tseng, J.-S. (2013). The impact of college students' intrinsic and extrinsic motivation on continuance intention to use English mobile learning systems. *The Asia-Pacific Education Researcher*, *22*(2), 181–192. https://doi.org/10.1007/s40299-012-0011-7

Chen, B., Wang, Y., & Wang, L. (2022). The effects of virtual reality-assisted language learning: A meta-analysis. *Sustainability, 14*(6), Article 3147. https://doi.org/10.3390/su14063147

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511486999

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Ding M. (2024). The impact of high-immersion virtual reality on EFL learners' foreign language speaking anxiety: A mixed-method approach. *ReCALL, 36*(3):287–305. https://doi.org/10.1017/S0958344024000156

Dionisio, J. D. N., Burns, W. G., & Gilbert, R. (2013). 3D virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys*, *45*(3), 1–38. https://doi.org/10.1145/2480741.2480751

Dolgunsöz, E., Yıldırım, G., & Yıldırım, S. (2018). The effect of virtual reality on EFL writing performance. *Journal of Language and Linguistic Studies, 14*(1), 278–292. http://www.jlls.org/index.php/jlls/article/view/869

Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning,* 1975-2000. https://doi.org/10.1080/09588221.2020.1854311

Godefridi, I., Suñer, F., Leblanc, C., & Meunier, F. (2021). Using virtual reality and peer feedback to reduce L2 speaking anxiety: An exploratory study. In N. Zoghlami, C. Brudermann, C. Sarré, M. Grosbois, L. Bradley, & S. Thouësny (Eds.), *CALL and professionalisation: Short papers from EUROCALL 2021* (pp. 100–105). Research-publishing.net. https://doi.org/10.14705/rpnet.2021.54.1316

Hamad, A., & Jia, B. (2022). How virtual reality technology has changed our lives: An overview of the current and potential applications and limitations. *International Journal of Environmental Research and Public Health, 19*(18), 11278. https://doi.org/10.3390/ijerph191811278

He, D. (2018). *Foreign language learning anxiety in China: Theories and applications in English language teaching*. Springer. https://doi.org/10.1007/978-981-10-7662-6

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLOS ONE, 13*(10), e0204494. https://doi.org/10.1371/journal.pone.0204494

Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal, 96*(2), 170–189. https://doi.org/10.1111/j.1540-4781.2011.01174.x

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hu-Au, E., & Lee, J. J. (2017). Virtual reality in education: A tool for learning in the experience age. *International Journal of Innovation in Education, 4*(4), 215–226. https://doi.org/10.1504/IJIIE.2017.10012691

Jenkins, J. (2009). *World Englishes: A resource book for students* (2nd ed.). Routledge.

Kachru, Y., & Smith, L. E. (2008). *Cultures, contexts, and world Englishes.* Routledge.

Kaplan-Rakowski, R., & Wojdynski, T. (2018). Students' attitudes toward high-immersion virtual reality assisted language learning. In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouësny (Eds.), *Future-proof CALL: Language learning as exploration and encounters – Short papers from EUROCALL 2018* (pp. 124–129). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.824

Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the Innovation in Language Learning International Conference 2019* (pp. 552–555). Pixel. https://conference.pixel-online.net/files/ict4ll/ed0012/FP/6143-SLA4232-FP-ICT4LL12.pdf

Kaplan-Rakowski, R., & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety. *Smart Learning Environments, 10*(1), Article 46. https://doi.org/10.1186/s40561-023-00263-9

Kawasumi, S., & Ishii, Y. (2023). Comparative study of English learning using virtual reality and a smartphone application. *Bulletin of the Faculty of Education, Chiba University, 71*, 99–105. https://doi.org/10.20776/S13482084-71-P99

Lin, T.-J., & Lan, Y.-J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational Technology & Society, 18*(4), 486–497. https://www.jstor.org/stable/jeductechsoci.18.4.486

Liu, X., & Hou, S. (2020). Effects of VR on users' speaking performance and cognitive load: Design of a VR-assisted English-speaking application on mobile phones. In *Proceedings of the 2020 International Conference on Intelligent Design (ICID)* (pp. 128–131). IEEE. https://doi.org/10.1109/ICID52250.2020.00035

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24–45). McGraw-Hill.

MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of foreign language anxiety: A review of the literature. *Language Learning, 41*(1), 85–117. https://doi.org/10.1111/j.1467-1770.1991.tb00677.x

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development, 66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Ministry of Education, Culture, Sports, Science and Technology. (2024). *Results of the FY2023 English education implementation survey* [令和5年度英語教育実施状況調査結果]. https://www.mext.go.jp/content/20240527-mxt_kyoiku01-000035833_1.pdf

Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a public university, Malaysia. *Teaching Public Administration, 43*(1), 4–20. https://doi.org/10.1177/01447394241230152

Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *TESOL Quarterly, 37*(4), 589–613. https://doi.org/10.2307/3588214

Park, S., Carlisle, D. D. E., Cheng, Z., Gillies, M., & Pan, X. (2025). Reducing foreign language anxiety through repeated exposure to a customizable VR public speaking application. *Frontiers in Virtual Reality, 6.* https://doi.org/10.3389/frvir.2025.1519409

Park, H. I., & Lee, A. R. (2005). L2 learners' anxiety, self-confidence and oral performance. In *Proceedings of the 10th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 197–208).

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology, Article 14.* https://doi.org/10.3389/fpsyg.2023.1255594

Qiu, X., Shan, C., Yao, J., & Fu, Q. (2024). The effects of virtual reality on EFL learning: A meta-analysis. *Education and Information Technologies, 29*(2), 1379–1405. https://doi.org/10.1007/s10639-023-11738-0

Schwienhorst, K. (2002). The state of VR: A meta-analysis of virtual reality tools in second language acquisition. *Computer Assisted Language Learning, 15*(3), 221–239. https://doi.org/10.1076/call.15.3.221.8186

Scrivner, O., Madewell, J., Buckley, C., & Perez, N. (2019). Best practices in the use of augmented and virtual reality technologies for SLA: Design, implementation, and feedback. In M. L. Carrió-Pastor (Ed.), *Teaching language and teaching literature in virtual environments* (pp. 55–72). Springer. https://doi.org/10.1007/978-981-13-1358-5_4

Thompson, A. S., & Lee, J. (2012). Anxiety and EFL: Does multilingualism matter? *International Journal of Bilingual Education and Bilingualism, 16*(6), 730–749. https://doi.org/10.1080/13670050.2012.713322

Thrasher, T. (2022). The impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility. *CALICO Journal, 39*(2), 219–238. https://doi.org/10.1558/cj.42198

Thrasher, T., Kaplan-Rakowski, R., Ovsiannikova, U., Meyr, J., & Yuan, Y. (2024). "I can talk to Spanish speakers in Illinois!": Student perspectives on AI-avatar role plays in virtual reality. In *WorldCALL2023: Conference proceedings* (pp. 157–165). IAFOR. https://doi.org/10.22492/issn.2759-1182.2023.18

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. https://doi.org/10.5539/ies.v2n4p39

Üstünbaş, Ü. (2024). Hey, GPT, Can we have a chat?: A case study on EFL learners' AI speaking practice. *International Journal of Modern Education Studies, 8*(1), 91–107. https://doi.org/10.51383/ijonmes.2024.318

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology, 14,* Article 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning, 54*(1), 119–152. https://doi.org/10.1111/j.1467-9922.2004.00250.x

Yıldız, C. (2024). ChatGPT integration in EFL education: A path to improved speaking self-efficacy. *Novitas-ROYAL (Research on Youth and Language), 18*(2), 167–182. https://doi.org/10.5281/zenodo.13861137

York, J., Shibata, K., Tokutake, H., & Koyama, T. (2021). Effect of SCMC on foreign language anxiety and learning experience: A comparison of voice, video, and VR-based oral interaction. *ReCALL, 33*(1), 49–70. https://doi.org/10.1017/S0958344020000154

Yudintseva, A. (2023). Virtual reality affordances for oral communication in English as a second language classroom: A literature review. *Computers & Education: X Reality, 2,* 100018. https://doi.org/10.1016/j.cexr.2023.100018

Appendix A

Student F's pre-transcript

Jamie: What's up?
Kei: Hi, I'm Kei Sato, please tell me your name.
Jamie: Hey, I'm Jamie. Nice to meet you, Kei.
Kei: Nice to meet you, too. Ah… Where, Where is your live?
Kei: Nice to meet you too. And where are you? Where do you live?
Jamie: I live in the city, not far from here. How about you, Kei?
Kei: I'm from Kanagawa in Japan.
Jamie: That's cool! I've always wanted to visit Japan. What's it like in Kanagawa?
Kei: I like that Kanagawa has a lot of delicious food and I hope you will come Kanagawa.
Jamie: I love trying food from different cultures! What's your favorite dish from Kanagawa?
Kei: My favorite dish from Kanagawa is Chinese food and how about you?
Jamie: I'm a big fan of sushi! The fresh fish and rice are just perfect together.
Kei: I hope you will come Japan another day.
Jamie: I'd love to visit Japan someday, especially to try the amazing food!
Kei: How about your favorite food in your city?
Jamie: I really enjoy the diverse food scene in my city, but I have a soft spot for the local barbecue joints. The smoky flavors are just irresistible!
Kei: That's nice and when you are see?
Jamie: Yeah. It's one of the things I love about living here.
Kei: I hope that I will see you tomorrow.
Kei: I will see you tomorrow.
Jamie: Looking forward to it! See you tomorrow, Kei.
Kei: And how do you feel?
Jamie: I'm feeling great, thanks for asking! See you tomorrow!

Student F's post-transcript

Jamie: What's up?
Kei: Hi, nice to meet you. What's your name?
Jamie: Hey there! I'm Jamie. What's yours?
Kei: I'm Kei. How are you?
Jamie: Hey Kei. I'm doing great, thanks. How about you?
Kei: I'm good, thanks. And where are live?
Jamie: Oh really?
Kei: I'm good, thanks. And where are live?
Jamie: I'm at The Deep Dive bar. It's a cool spot. Have you been here before?
Kei: No, I haven't. What favorite your drink?
Jamie: I love a good mojito, especially on a warm day. What about you?
Kei: I love beer and
Kei: I like beer and
Kei: I like a beer.

Jamie: Nice choice! Sometimes a cold beer is just what you need. So, did you play any instruments?
Kei: Sorry, what do you mean, instrument?
Jamie: An instrument is something you use to make music, like a guitar or a piano. Do you play any music?
Kei: Yes, I play the guitar. How about you?
Kei: Yes, I play the guitar. How about you?
Kei: Yes, I play the guitar. How about you?
Jamie: That's awesome! I also play the guitar. It's my favorite instrument. Maybe we can jam together sometime.
Kei: Yeah, I want to see you again.
Jamie: Sure, that would be great! Let's plan to meet here again. Looking forward to it!

Note: Jamie is an avatar's name. For privacy protection, "Kei" is a pseudonym used in both of the transcripts. In the post section, Kei sounded uncertain and hesitated while speaking the line "Yes, I play the guitar. How about you?" and repeated it.

Appendix B

The Foreign Language Anxiety questionnaire

1. **I never feel quite sure of myself when I am speaking in my foreign language class.**
2. I don't worry about making mistakes in language class.
3. **I tremble when I know that I'm going to be called on in language class.**
4. **It frightens me when I don't understand what the teacher is saying in the foreign language.**
5. It wouldn't bother me at all to take more foreign language classes.
6. **During language class, I find myself thinking about things that have nothing to do with the course.**
7. **I keep thinking that the other students are better at languages than I am.**
8. I am usually at ease during tests in my language class.
9. **I start to panic when I have to speak without preparation in language class.**
10. **I worry about the consequences of failing my foreign language class.**
11. I don't understand why some people get so upset over foreign language classes.
12. **In language class, I can get so nervous I forget things I know.**
13. **It embarrasses me to volunteer answers in my language class.**
14. I would not be nervous speaking the foreign language with native speakers.
15. **I get upset when I don't understand what the teacher is correcting.**
16. **Even if I am well prepared for language class, I feel anxious about it.**

17. **I often feel like not going to my language class.**
18. I feel confident when I speak in foreign language class.
19. **I am afraid that my language teacher is ready to correct every mistake I make.**
20. **I can feel my heart pounding when I'm going to be called on in language class.**
21. **The more I study for a language test, the more confused I get.**
22. I don't feel pressure to prepare very well for language class.
23. **I always feel that the other students speak the foreign language better than I do.**
24. **I feel very self-conscious about speaking the foreign language in front of other students.**
25. **Language class moves so quickly I worry about getting left behind.**
26. **I feel more tense and nervous in my language class than in my other classes.**
27. **I get nervous and confused when I am speaking in my language class.**
28. When I'm on my way to language class, I feel very sure and relaxed.
29. **I get nervous when I don't understand every word the language teacher says.**
30. **I feel overwhelmed by the number of rules you have to learn to speak a foreign language.**
31. **I am afraid that the other students will laugh at me when I speak the foreign language.**
32. I would probably feel comfortable around native speakers of the foreign language.
33. **I get nervous when the language teacher asks questions which I haven't prepared in advance.**

Note. The items used in this study are bold.

Appendix C

1. I can learn English at any time with Immerse.
2. I can learn English at any place with Immerse.
3. Immerse is convenient for me to engage in practicing speaking English.
4. I feel that Immerse is convenient for me to practice speaking English.
5. When I learn English with Immerse, time passes quickly.
6. When I learn English with Immerse, I am not distracted by noise.
7. Practicing speaking English via Immerse makes me feel good.
8. Practicing speaking English via Immerse is enjoyable to me.
9. Learning to operate Immerse is easy for me.
10. My interaction with Immerse is clear and understandable.
11. It is easy for me to become skillful at using Immerse.
12. I find Immerse easy to use.
13. Using Immerse enables me to improve my English-speaking skills.

14. Using Immerse makes it easier for me to practice speaking English.
15. I find Immerse useful for practicing speaking English.
16. I would like to continue to practice speaking English using Immerse in the future.
17. I plan to practice speaking English using Immerse in the future.
18. I expect my use of Immerse to continue in the future.