

THE EFFECT OF IMMERSIVE CONVERSATIONAL AI ON IMMIGRANT STUDENTS'  
SECOND LANGUAGE PRODUCTION AND WILLINGNESS TO COMMUNICATE

BY

ULIANA OVSIANNIKOVA

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Arts in Teaching of English as a Second Language  
in the Graduate College of the  
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Master's Committee:

Professor Randall Sadler, Chair  
Dr. Tricia Thrasher, Director of Research, Immerse  
Assistant Professor Aurore Mroz  
Professor Emerita Dorothy Chun, University of California, Santa Barbara

## ABSTRACT

Recognizing the challenges that immigrant Latino youth in the U.S. encounter, there is a growing interest in harnessing Artificial Intelligence (AI) for English as a Second Language (ESL) learning (Ayedoun et al., 2019). While existing research predominantly involves university-level participants (Woo & Choi, 2021) and focuses on students' perceptions (Jeon & Choe, 2023), there is a need to investigate the potential of conversational AI for ESL in a naturalistic high-school setting (Katsarou et al., 2023). Moreover, there has been limited research on immersive conversational AI, which combines the benefits of Virtual Reality (VR) in reducing the affective filter (Thrasher, 2022) with AI's potential for personalized learning experiences (Huang et al., 2023).

Responding to the call for research on immersive conversational AI, this mixed-methods study, guided by the ecological perspective and Complex Dynamic Systems Theory (CDST) (Larsen-Freeman, 2018), explores the impact of immersive conversational AI on language production and willingness to communicate (WTC) in an underserved Illinois high school. Over one academic quarter, thirty-six ( $N = 36$ ) immigrant Latino high school students in ESL Resources class engaged in 20-minute-long weekly role-plays with AI-powered bots through the *Immerse* application, which was first accessed on desktop and subsequently in VR. Interactions with AI bots were recorded and analyzed for students' language production and usage of conversational features. Students' WTC was also evaluated through a questionnaire by Dewaele and Dewaele (2017). During data analysis, a Spearman correlation test was conducted to explore relationships between ESL level, immersion level, modality preference, WTC, language production, and specific conversational features.

Findings from this study indicate that lower-level ESL students preferred desktop interactions and demonstrated lower initial willingness to communicate, but they exhibited significant growth in WTC compared to advanced ESL peers, who favored VR and started with higher WTC levels. Beginner and low-intermediate students using desktop showed greater growth in language production. While lower-level students engaged more frequently with all conversational features, only the word pronunciation feature positively correlated with higher language production. Additionally, two disruptions in human-AI communication patterns were identified: limited recognition of mispronounced words, impacting lower-level ESL students, and issues with recognizing contextual cues across all levels. These findings underscore conversational AI's potential to enhance language production and willingness to communicate among lower-level immigrant ESL students, supporting its integration into high school curricula.

## ACKNOWLEDGMENTS

Pursuing my MATESL degree at the University of Illinois Urbana-Champaign has been a transformative journey that expanded my passion for teaching into a deep interest in research and educational technologies. I am profoundly grateful to my advisor, Dr. Randall Sadler, whose unwavering support and guidance have been essential throughout these two years. Working as your research assistant and integrating VR into high school language programs was a critical step in my understanding of immersive technologies which inspired my research interest. Having you as my advisor has profoundly impacted my academic and professional path, for which I am deeply appreciative.

This study would not have been possible without the invaluable contributions of my thesis committee members. I extend my sincere appreciation to Dr. Aurore Mroz for her dedication and insightful feedback. Your guidance transformed the initial overwhelming exploration of this novel technology into a coherent and structured study. I am particularly grateful for your detailed feedback and the knowledge I have about mixed-methods research.

My deepest gratitude also goes to Dr. Tricia Thrasher, whose constant support and high standards of work ethic have been a source of inspiration for me. Working with you on the VR research project has been an incredible and enjoyable learning experience. Your mentorship extended beyond academic guidance, helping me grow both professionally and personally. Thank you for your tireless efforts in training participants to use VR and helping me organize this study.

I am also deeply appreciative of Dr. Dorothy Chun's guidance. Your insights into qualitative research pushed me to delve deeper and uncover the core affordances of immersive

technologies. Your encouragement to critically and thoroughly analyze data has been instrumental in enhancing the quality and depth of this study.

Finally, I would like to express my profound gratitude to my family for their unwavering support throughout my academic journey across five countries. Your belief in me and encouragement have always given me strength to follow my dreams. To my friends, thank you for inspiring me with your achievements and motivating me to strive for excellence. And to my wonderful partner, Radule, your patience, love, and support, even from afar, have been my anchor. Thank you for being my home away from home and always taking care of me.

## TABLE OF CONTENTS

<b>LIST OF TABLES .....</b>	<b>viii</b>
<b>LIST OF FIGURES .....</b>	<b>ix</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>3</b>
THEORETICAL FRAMEWORK .....	3
LINEAGE AND GAP IN RESEARCH.....	6
RATIONALE FOR MIXED METHODS .....	10
RESEARCH QUESTIONS .....	11
<b>CHAPTER 3: METHODOLOGY .....</b>	<b>12</b>
OVERVIEW OF DESIGN .....	12
PARTICIPANTS.....	17
DATA COLLECTION .....	19
DATA ANALYSIS .....	22
<b>CHAPTER 4: FINDINGS .....</b>	<b>30</b>
DESKTOP PREFERENCE OVER VR HEADSETS .....	30
RQ1: IMPACT OF AI ON WTC .....	30
RQ2: IMPACT OF AI ON LANGUAGE PRODUCTION .....	37
RQ3: RECURRING PATTERNS IN HUMAN-AI COMMUNICATION BREAKDOWNS .....	42
<b>CHAPTER 5: DISCUSSION .....</b>	<b>50</b>
PREFERRED LEVEL OF IMMERSION.....	51

WILLINGNESS TO COMMUNICATE .....	53
LANGUAGE PRODUCTION .....	54
INDIVIDUAL DIFFERENCES IN HUMAN-AI CONVERSATIONS .....	57
<b>CHAPTER 6: CONCLUSION.....</b>	<b>59</b>
LIMITATIONS.....	60
IMPLICATIONS .....	61
FUTURE RESEARCH.....	62
<b>REFERENCES.....</b>	<b>63</b>
<b>APPENDIX A: INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL .....</b>	<b>72</b>
<b>APPENDIX B: LANGUAGE AND SOCIAL BACKGROUND QUESTIONNAIRE .....</b>	<b>73</b>
<b>APPENDIX C: WILLINGNESS TO COMMUNICATE QUESTIONNAIRE .....</b>	<b>83</b>
<b>APPENDIX D: PYTHON CODES FOR SPEARMAN CORRELATIONS .....</b>	<b>86</b>

## LIST OF TABLES

TABLE 1: PARTICIPANTS' GENDER AND GRADE .....	19
TABLE 2: DATA COLLECTION TIMELINE .....	20
TABLE 3: ESL LEVEL AND STUDENTS VR USAGE DURING THE STUDY .....	30
TABLE 4: WTC GROWTH BASED ON LEVEL OF IMMERSION.....	31
TABLE 5: RECURRING PATTERNS IN COMMUNICATION DISRUPTION .....	43

## LIST OF FIGURES

FIGURE 1: OVERVIEW OF STUDY DESIGN.....	13
FIGURE 2: IMMERSE LIVE TEACHER-LED LESSON OVERVIEW.....	14
FIGURE 3: META QUEST 2 VR HEADSET.....	15
FIGURE 4: IMMERSE AI INTERFACE.....	16
FIGURE 5: RESEARCHER’S POSITIONING STATEMENT.....	17
FIGURE 6: STUDENT TRAINING USING META QUEST 2 VR HEADSETS.....	21
FIGURE 7: CONVERSATIONAL FEATURE "HINTS".....	26
FIGURE 8: CONVERSATIONAL FEATURES "WORD DEFINITION" AND "PRONUNCIATION".....	27
FIGURE 9: WTC SCORES FOR DESKTOP AND VR GROUPS.....	33
FIGURE 10: PERCENTAGE GROWTH IN WTC BASED ON ESL LEVEL.....	34
FIGURE 11: SPEARMAN CORRELATION WTC, ESL LEVEL, AND LEVEL OF IMMERSION.....	35
FIGURE 12: SPEARMAN CORRELATION MATRIX FOR PERCENTAGE GROWTH IN LANGUAGE PRODUCTION.....	38
FIGURE 13: SPEARMAN CORRELATION MATRIX FOR CONVERSATIONAL FEATURES AND LANGUAGE PRODUCTION.....	40
FIGURE 14: SPEARMAN CORRELATION MATRIX FOR LANGUAGE PRODUCTION AND FEATURES WITH P-VALUE < .05.....	41
FIGURE 15: LIMITED RECOGNITION OF MISPRONOUNCED WORDS.....	44
FIGURE 16: LIMITED RECOGNITION OF CONTEXTUAL CUES, ESL 2 EXAMPLE.....	46
FIGURE 17: LIMITED RECOGNITION OF CONTEXTUAL CUES, ESL 4 EXAMPLE.....	47

## CHAPTER 1: INTRODUCTION

Immigrant Latino youth in the U.S. face diverse challenges in their process of adaptation in the society, with limited English proficiency being among the key obstacles to their successful integration in the mainstream classrooms (Lu et al., 2021; Ortiz et al., 2019). Refugee students with limited or interrupted formal education (SLIFE) struggle especially with building their proficiency in a second language, which affects their academic success (DeCapua, 2011; Li, 2017) and cross-cultural adaptation (Gallagher, 2013). In the face of these challenges, Computer Assisted Language Learning (CALL) in general, and Artificial Intelligence (AI) in particular, have attracted scholars' attention for their potential to beneficially contribute to second language acquisition (SLA), including second language production and willingness to communicate (Ayedoun et al., 2019). However, most AI studies have focused on the university setting (Huang et al., 2023; Woo & Choi, 2021) leaving immigrant high school students underrepresented in the CALL research. Moreover, with further advances in the technological capabilities of AI, the potential of immersive conversational AI, which combines immersive experience of Virtual Reality (VR) with AI's advanced Natural Language Processing (NLP) capabilities, should be explored more in the realm of SLA (Davis, 2022; Katsarou et al., 2023; Ruan et al., 2021). While existing studies on immersive conversational AI were primarily quantitative or mixed methods with the qualitative strand represented solely by open-ended questions in questionnaires (Jeon & Choe, 2023; Thrasher et al., 2024), a mixed-methods study design with dominating qualitative strand is needed to provide a deeper understanding of this technology for language production in written and spoken modalities in a foreign language.

To grasp the holistic impact of conversational AI on language production and willingness to communicate, this current study was conducted as an extension of a large-scale study of

Thrasher et al. (2024) on student perception of immersive conversational AI. This current study adopts an embedded mixed-methods design for the purpose of complementarity, where quantitative and qualitative data is analyzed in a way that "one data set provides a supportive, secondary role" (Creswell & Plano Clark, 2011, p. 91). Guided by the ecological perspective and Complex Dynamic Systems Theory (CDST) (Larsen-Freeman, 2018), this study primarily relies on qualitative data in a form of recordings of students' interactions with conversational AI to explore the development of language production in human-AI conversations. The study secondarily relies on quantitative data from questionnaires to complementarily analyze impact of conversational AI on willingness to communicate in English as a Second Language (ESL).

## CHAPTER 2: LITERATURE REVIEW

### Theoretical Framework

The current study follows the ecological perspective, where “SLA does not take place in static isolation” but is “emergent from and dynamically interconnected with the environment.” (Larsen-Freeman, 2018, p. 59). Ecological orientation focuses on the inherent connection that exists between the learner and different variables, such as space, tools, and social interactions (Kramsch & Steffensen, 2008). Researchers emphasized the need for ecological perspective in CALL research, which tends to be too focused on technocentric assumptions but instead needs to account for the complex dynamic of social interactions among learners with the environment and tools (Mroz, 2015).

This study is framed by Complex Dynamic Systems Theory (CDST) which conceives of second language acquisition as a transforming dynamic system. According to CDST, "language development occurs at the nexus, or intersection, of [cognitive involvement and social interaction]" (Larsen-Freeman, 2018, p. 58). CDST, following the ecological perspective, accounts for the individual differences in the learning process and has been used in the research on emotions in L2 acquisition (Thrasher, 2022).

#### *AI and VR*

Artificial Intelligence (AI), a technology capable of performing tasks that require human intelligence (Bolter, 1984), has been an active area of study and development for the past 70 years. Generative AI, a more recent subfield, focuses on creating models that generate new data resembling provided input data (Goodfellow et al., 2014). Recent advancements in AI, including content creation, speech recognition, and natural language processing, have led to the emergence

of Conversational AI. This technology simulates human interaction and has been actively studied in the field of Second Language Acquisition (SLA) (Davis, 2022; Jeon & Choe, 2023). This study explores students' interactions with conversational AI, powered by the GPT-3 language model of OpenAI.

Virtual Reality (VR) is defined as a technology that creates immersive experiences through high-level interactions (Dwivedi et al., 2022). Based on the intensity, there are two levels of immersion defined by Kaplan-Rakowski and Gruber (2019): low-immersion and high-immersion VR. Low-immersion VR, also called "desktop VR," can be exemplified by *Second Life*, where interaction with the 3D environment usually occurs on a desktop monitor (Sadler, 2020). High-immersion VR creates a higher sense of authenticity and "being there" and can be defined as "a computer-generated 360° virtual space that can be perceived as being spatially realistic due to the high immersion afforded by a head-mounted device" (Kaplan-Rakowski & Gruber, 2019, p. 552). In this study, conversations with AI bots take place in immersive environments accessible at both levels of immersion. Participants first had conversations in a low-immersion environment, referred to as "desktop," and after three sessions, they could use converse in high-immersion environment, referred to as "VR" in this study.

The combination of virtual reality and conversational AI leads to the emergence of immersive conversational AI, creating highly authentic immersive experiences simulating human conversation. According to Jeon and Choe (2023), conversational AI bots have a specific combination of three main affordances: goal-orientation, multimodality, and embodiment. While some chatbots are highly focused on language learning and support only written modality, immersive conversational AI exhibits a combination of these three affordances: goal-orientation (as bots are programmed according to CEFR levels), multimodality of both output (seeing text

and hearing it) and input (speaking and writing), and embodiment with a 3D avatar in an authentic environment where dialogue takes place.

### *Willingness to Communicate and Language Production*

The established theoretical framework is highly suited for the study of willingness to communicate (WTC) which can be defined as “a readiness to initiate discourse with specific person(s) at a particular time” (MacIntyre et al., 1998, p. 547). A substantial number of researchers addressed the dynamic complexity of individual willingness to communicate. In particular, Lee and Lee (2020) highlighted the interconnection between “individual trait-like and contextual state-like variables” in developing individual willingness to communicate. While most studies investigate the development of WTC from the spoken language perspective, there have been a limited number of studies addressing WTC in written communication (Tai, 2017), which has been growing in importance in both human-to-human communication and especially in human-computer interaction.

Language production in SLA refers to the ability to produce speech or writing in a nonnative language (Gass & Varonis, 1994). It involves the formulation of the intended message and its production using linguistic elements. As “language production is a form of action” (Pickering & Garrod, 2013, p. 337), numerous interconnected variables, both internal and external, can affect language production at any given moment. Therefore, a combination of multiple variables needs to be considered while assessing language production, especially in a foreign language.

As this study is framed by Complex Dynamic Systems Theory (CDST) and the ecological perspective, the process of second language acquisition is viewed as a dynamic system unique to each individual. Therefore, the study does not compare students’ language

production and willingness to communicate scores across different proficiency levels. Instead, it explores students' individual progress in these variables and communication patterns with AI that influence these variables.

### **Lineage and Gap in Research**

According to Edwards et al. (2014) and Lu et al. (2021), immigrant Latino youth are exposed to various stressors on a daily basis which affects their motivation, self-confidence, and anxiety. Due to accumulated stress caused by traumatic experience and difficult social adaptation (Alefesha & Al-Jamal, 2019; Gallagher, 2013; Li & Grineva, 2017), immigrant high school students may have a predisposition to a high affective filter which hinders their second language acquisition (SLA) and affects their willingness to communicate (WTC) in a foreign language (Krashen, 1982). Taking into account the mixed-ability nature of ESL refugee classes and the various stressors immigrant Latino youth face during the adaptation process, DeCapua (2011) emphasizes the need for effective language acquisition strategies, which would provide sufficient support to ESL students.

Considering the affordances of AI technology in "providing personalized learning experiences" and "enabling immediate adjustment... after receiving automated feedback" (Huang et al., 2023, p.126), AI has the potential to support both immigrant ESL students and instructors by reducing challenges of high school students in gaining oral proficiency in ESL. Researchers in the field of CALL have been exploring the affordances of AI for second language acquisition, notably its impact on language production (Hsu et al., 2021; Moussali & Cardoso, 2020) and willingness to communicate (Ayedoun et al., 2019; Tai & Chen, 2020).

## *Research on AI for language learning*

Studies on artificial intelligence for language learning have primarily focused on intelligent tutoring systems (ITSs) (Chen et al., 2022), which provide personalized feedback but have a limited set of responses to simulate human-like conversation (Ruan et al., 2021). Even though AI developers and researchers have noted the technology's affordances for speaking skills (Woo & Choi, 2021), more than 50 % of publications have been devoted to using intelligent tutoring systems (ITSs) for writing, and only 6 % have focused on ITSs for speaking. This is based on a bibliometric analysis of 516 articles on artificial intelligence for language learning from 2000 to 2019 (Huang et al., 2023). It can be explained by the prevalence of university settings in research on both immigrants' second language acquisition (Kanno & Varghese, 2010) and AI for language learning (Woo & Choi, 2021), as writing skills are identified as the major struggle for ESL university-level students (Kanno & Varghese, 2010). Notably, little research has been conducted on AI for language learning with high school participants. Only one out of 53 studies in a systematic literature review on AI in language education focused on high school students (Woo & Choi, 2021).

With the development of speech recognition and natural language processing (NLP), new generations of conversational AI or dialogue-based AI have the potential to improve students' speaking skills as well (Davis, 2022). According to two meta-analyses by Katsarou et al. (2023) and Woo and Choi (2021), intelligent virtual assistants (IVAs), such as *Alexa* and *Siri*, have been studied the most to determine their positive effects on oral production (Hsu et al., 2021; Tai, 2022) and willingness to communicate (Tai & Chen, 2020). While the study by Dizon and Tang (2020) claims that this technology has been viewed positively by students, improving autonomous learning and vocabulary acquisition, there are significant gaps in the available data.

Although 20 participants were initially recruited, six did not start the study, three stopped using it after one try, and only four participants used *Alexa* for more than 16 days out of a six-month-long study. According to Tai (2022), 89 university-level students who used *Google Assistant* achieved a comparable level of oral proficiency to those conversing with a native English speaker. Finally, in their short-term experimental study on using *Google Assistant*, Tai and Chen (2020) found that this technology led to increased willingness to communicate among high school students, lower foreign language anxiety, and increased engagement.

According to the Jeon and Choe (2023) framework for defining conversational AI based on three key affordances, immersive conversational AI encompasses signs of goal orientation, embodiment, and multimodality. However, research on high-immersion conversational AI with more advanced embodiment is very limited compared to the low-immersion kind. Studies on conversational AI chatbots in low-immersion VR with text, voice, and a 2D image have significantly higher numbers of participants (Ayedoun et al., 2019), longer durations (Wang et al., 2022), and a higher diversity of educational settings (Yang et al., 2022) than the limited number of studies on conversational AI in high-immersion VR. According to Ayedoun et al. (2019), a study conducted on 40 ESL university-level participants found that a chatbot using both communication strategies and affective backchannels—responses indicating encouragement and emotional support—improved second language learners' willingness to communicate more than a chatbot using only one of these features. Studies by Yang et al. (2022) and Wang et al. (2022), both conducted with over 300 elementary school participants, found a connection between students' positive perceptions of the technology and the human-likeness of the conversational AI, which further led to higher motivation.

### *Immersive Conversational AI*

Guided by the incentive to provide learners with an even more human-like conversational partner to practice production skills, more researchers have turned their attention to immersive conversational AI, combining the affordances of AI with the benefits of VR (Divekar et al., 2022). Surprisingly, the presented benefits of conversational AI highly correlate with findings from VR studies. In particular, VR has a positive effect on oral production (Dooly et al., 2023; Thrasher, 2022), foreign language anxiety (Kaplan-Rakowski & Gruber, 2023; Thrasher, 2022; Xie et al., 2019), improved pronunciation (Thrasher, 2022), motivation (Xie et al., 2019), and higher engagement (Kaplan-Rakowski & Gruber, 2023). Combining the affordances of conversational AI and high-immersion VR justifies the increased research interest in immersive conversational AI. However, due to the novel nature of this technology, the few studies that exist have primarily explored the perception and usability of immersive conversational AI rather than actual learning gains (Mirzaei et al., 2018). In particular, a large-scale study on integrating VR headsets in high schools and colleges in Illinois, Texas, and California included a sub-study specifically on students' perceptions of immersive conversational AI (Thrasher et al., 2024). According to Thrasher et al. (2024), students did not experience difficulty understanding AI bots and had an overall positive perception of the technology. However, the study solely relied on questionnaire responses and did not explore the effect of conversational AI on affective factors or language production in a second language.

Existing studies by Divekar et al. (2021) and Hassani et al. (2016) that combine the affordances of conversational AI and virtual reality are short-term and conducted in a lab setting: Divekar et al. (2021) on Chinese as a Foreign Language (CFL) learners and Hassani et al. (2016) on English as a Second Language (ESL) learners. Both of these studies, having a sample size of

10 university-level participants, are primarily quantitative with the only qualitative aspect in Divekar et al. (2021) presented through open-ended questionnaires. While Hassani et al. (2016) claimed significant learning gains in students' speaking skills, Divekar et al. (2021) expanded learning gains in vocabulary acquisition with findings on students' positive perception of the technology along with increased willingness to communicate. Even though learning gains after using immersive conversational AI included more accurate replies, fewer grammatical errors, shorter pronunciation duration, and overall proficiency (Hassani et al., 2016), both studies, conducted short-term in a non-naturalistic setting, have lower ecological validity, so their findings may not be replicable in a real-life classroom situation.

The shared limitations of these studies lead to the call for research on the development of second language production during the interaction with immersive conversational AI, especially from the ecological perspective, where SLA is viewed as a dynamic process with inherently interconnected variables (Larsen-Freeman, 2018).

### **Rationale for Mixed Methods**

Following the call for research, this study adopts an embedded mixed-methods design with a primary qualitative strand and complementary quantitative data. Even though a substantial number of studies on conversational AI have a mixed-methods design to complement a small sample size, they primarily rely on quantitative data in a form of questionnaires and a qualitative strand is usually presented by open-ended questions (Katsarou et al., 2023). While some mixed-methods studies primarily rely on qualitative data in a form of recordings or observations, it is rarely merged with quantitative data to provide a deeper understanding of how the affordances of the technologies contributed to growth in language production (Moussali & Cardoso, 2020). Moreover, continuous technological advancement of artificial intelligence leads to the increasing

number of studies with an exploratory design, where AI developers analyze students' perception of the novel human-like conversational AI for speaking practice with a limited focus on learning gains (Davis, 2022; Ruan et al., 2021). Hence there is a need for research which triangulates qualitative observations of human-AI interactions with quantitative data for better understanding of affordances of this technology for language learning.

### **Research Questions**

Based on the previous research, this study aims to address identified research gaps by answering the following research questions:

RQ1: What is the impact of conversational AI on students' willingness to communicate based on their ESL level and level of immersion?

RQ2: How does conversational AI affect second language production, based on the type of immersive environment, WTC, participants' ESL level, and communication modality?

RQ2.1: What features of conversational AI positively affect language production, and which features negatively affect language production?

RQ2.2: Is there a difference between the features used by high-level ESL students compared to low-level ESL students?

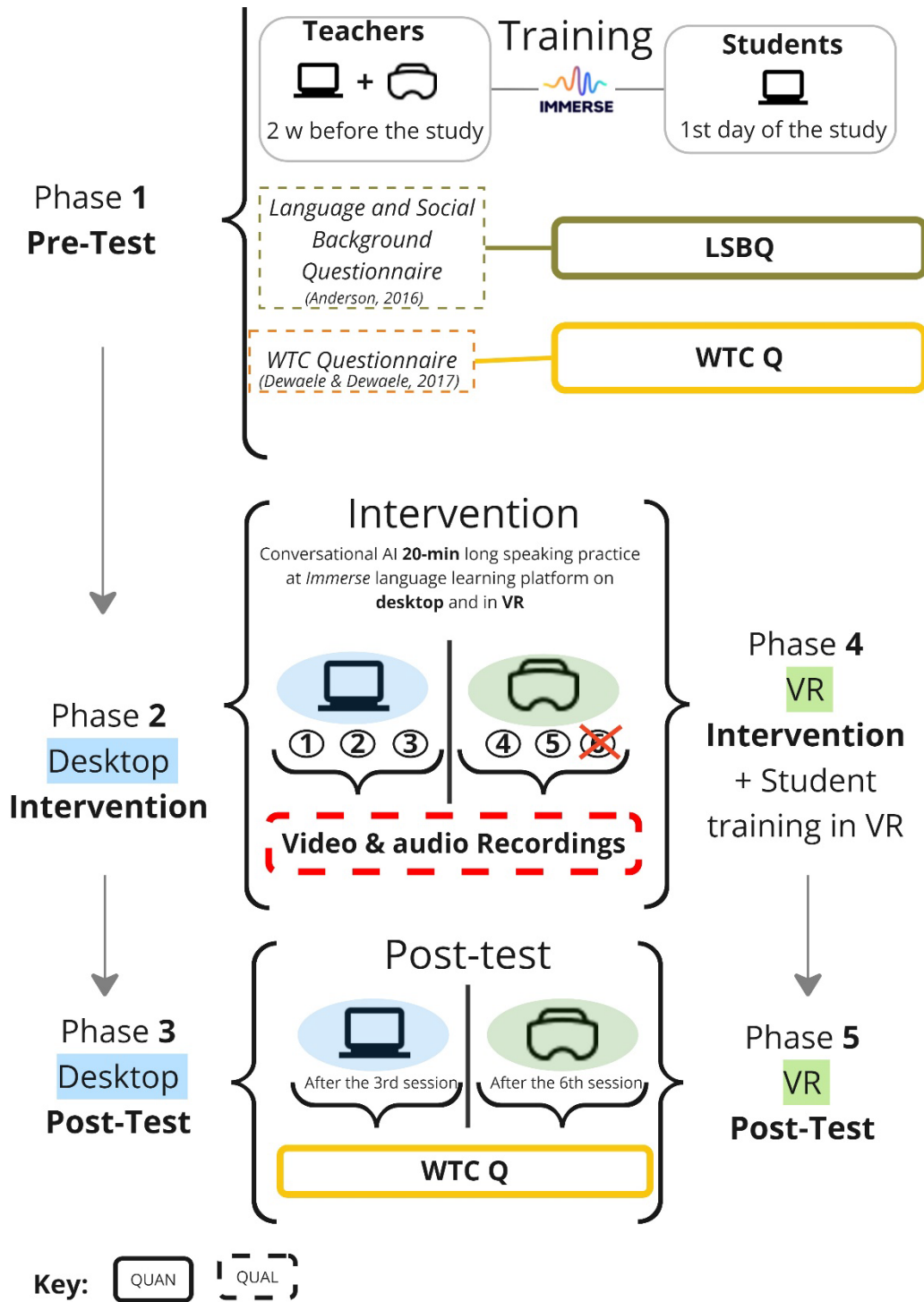
RQ3: What are the patterns and causes of communication disruptions in human-AI interactions?

## CHAPTER 3: METHODOLOGY

### Overview of Design

This mixed-method study follows Creswell and Plano Clark's (2011) *embedded design* which predominantly relies on qualitative data and integrates a quantitative strand to answer research questions which require a triangulation of quantitative and qualitative data. While one variable, namely willingness to communicate, fully depends on quantitative strand of WTC questionnaires, language production is primarily elicited in video recordings of students' interactions with conversational AI. The study follows five main phases: pre-test phase; two intervention phases, divided into low-immersion (desktop) sub-phase and high-immersion (VR) sub-phase, each entailing several points of data collection, as per CDST; and two post-test phases after each of the sub-phases of the intervention. Figure 1 presents an overview of the study design.

**Figure 1: Overview of Study Design**



Phase 1 pre-test aimed to identify the baseline of students' willingness to communicate in ESL along with their language and social background. Phase 1 started with teacher training to use *Meta Quest 2* VR headsets and *Immerse*, a virtual reality platform for immersive conversational AI used for this study, two weeks before the beginning of the study. *Immerse* primarily offers teacher-led live lessons in Spanish, French, and ESL from foundations to advanced levels, and can be accessed either via a VR headset or via desktop. It also offers complementary AI role plays in 9 immersive environments for asynchronous conversation practice, which were used in this current study. An overview of *Immerse* is presented in Figure 2.

**Figure 2:** Immerse live teacher-led lesson overview



Phase 1 also encompassed student training on using *Immerse* on *Chromebooks* on the first day of the study.

The next stage is Phase 2 desktop intervention, which encompassed the integration of conversational AI in a low-immersion environment (on participants' *Chromebooks*) for three weekly sessions. At the end of the third session, the next Phase 3 desktop post-test began. Phase

3 aimed to collect post-intervention data to serve as the final progress point of the identified variables for students' individual progress in a low-immersion environment.

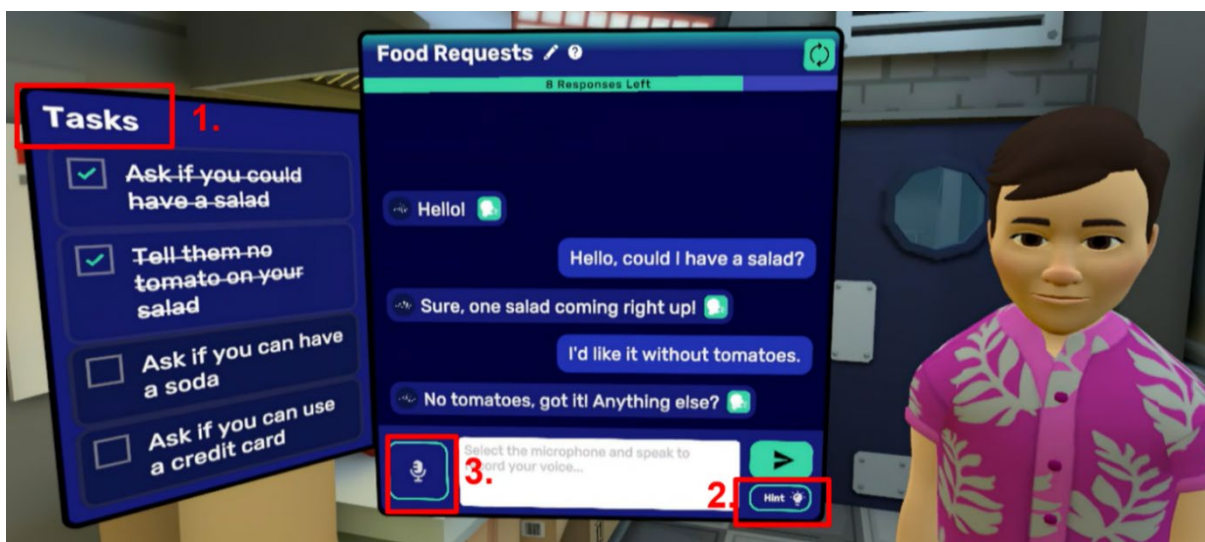
Phase 4 VR intervention replicated Phase 2 desktop intervention but in a high-immersion environment (i.e. VR) for the duration of two more sessions in VR that followed three sessions on desktop in Phase 2. *Meta Quest 2* VR headsets, presented in Figure 3, were used for this study. While the original study design had an equal distribution of low- and high-immersion sessions (i.e., three on desktop and three in VR), the last VR session was canceled due to teachers' overloaded schedules at the end of the academic quarter. Finally, during Phase 5 VR post-test, the same data as in the Phase 3 desktop post-test was collected.

**Figure 3:** Meta Quest 2 VR headset



A virtual reality language learning application, *Immerse*, was chosen as a platform where students interact with conversational AI for the following reasons. First, *Immerse* developed human-like conversational AI powered by ChatGPT in accordance with ACTFL proficiency guidelines to adapt the level of interaction difficulty to a student's proficiency level. Second, *Immerse* offers access to their platform in two types of immersion: low-immersion on desktop and high-immersion in VR, which enables the analysis of under-researched immersive conversational AI. Finally, the *Immerse* platform offers 90 possible conversations with AI in 9 immersive environments, such as fast-food restaurants, homes, garages, doctor's offices, and others. For example, Figure 4 shows a beginner-level conversation with the AI bot in a fast-food restaurant to practice the topic of food requests. In the figure, number one indicates tasks that the student needs to complete during the conversation. If the student needs help with completing the conversational tasks, they can click on the hint button, marked by number two. Finally, the student can record their response by clicking on the microphone icon to start speaking. There is also a chat dialogue available to enter their written response.

**Figure 4:** Immerse AI interface



Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig3>

The researcher's positioning statement is provided in Figure 5 to give more contextual information about this study.

**Figure 5:** Researcher's Positioning Statement

---

This study is situated within a large-scale study on integrating Virtual Reality into high school language classes. The large-scale study spans Illinois, Texas, and California, supported by a generous grant of 500 *Meta Quest 2* headsets, which were distributed to high schools, colleges, and universities across these three states. As a Research Assistant (RA) for the Illinois part of the project, I had the honor of working with Dr. Tricia Thrasher, Director of Research at Immerse, and Dr. Randall Sadler, PI for the Illinois side of the project. I was responsible for setting up the headsets, providing student and teacher training, and developing supporting materials and lesson plans in VR to teach Spanish and French on the *Immerse* platform using *Meta Quest 2* headsets.

Among the four participating schools, one underserved Illinois school agreed to participate in the study for my master's thesis. Even though I was able to access immigrant high school students and distribute VR headsets thanks to the large-scale project, this study primarily focuses on conversational AI, and was designed independently of the large-scale study.

---

### **Participants**

This study adopted a non-probabilistic convenience sampling and was conducted with 36 immigrant Latino high school students taking an ESL Resources class for the duration of one academic quarter (9 weeks) at an underserved high school in Illinois. International students willing to attend the high school are required to take an ESL placement test, which determines

students' levels from ESL 1 to ESL 4 and assigns the corresponding class the students are required to take for one academic quarter. Once the students take an ESL class, they are required to register for an ESL Resources class, which is the setting for this study. Compared to the traditional ESL 1 to ESL 4 classes, the ESL Resources class does not have a fixed curriculum, as it provides academic support for mainstream classes to ESL students and reinforces students' autonomy. Each 80-minute-long class takes place five times a week. Among the typical activities students do during the class are games in English, retaking tests from content classes, doing homework for content classes, and conversing with peers. As students are often not given any assignments during the class, they sometimes play games, follow social media, and use video-streaming platforms during the class time.

The research was conducted with the whole ESL department of the high school, consisting of three ESL high school teachers. For the duration of the Fall 2023 academic quarter, a 20-minute-long weekly practice with conversational AI was integrated into four ESL Resources classes. A total number of 36 ( $N = 36$ ) participants were recruited from four ESL Resources classes taught by three teachers. As these classes do not have a predefined curriculum, students taking this class exhibit varying levels of English proficiency: from beginners required to attend ESL 1 class to highly advanced students required to take ESL 4 class focusing on English writing composition. The participants' basic demographic information is presented in Table 1.

**Table 1:** Participants' Gender and Grade

Section	Participants						Total
	Male			Female			
	9th	10th	11th	9th	10th	11th	
1 <sup>st</sup> h. Instructor 1	0	2	1	1	0	2	<b>6</b>
2 <sup>nd</sup> h. Instructor 2	3	2	1	4	2	0	<b>12</b>
3 <sup>rd</sup> h. Instructor 3	4	2	0	1	0	1	<b>8</b>
4 <sup>th</sup> h. Instructor 3	7	0	0	0	2	1	<b>10</b>
<b>Total:</b>	<b>14</b>	<b>6</b>	<b>2</b>	<b>6</b>	<b>4</b>	<b>4</b>	<b>36</b>

The first class hour consisted of 6 participants ( $n = 6$ ) taught by Instructor 1, the second hour consisted of 12 participants ( $n = 12$ ) taught by Instructor 2, the third hour had 8 participants ( $n = 8$ ), and the fourth hour had 10 participants ( $n = 10$ ), both taught by Instructor 3. Across these sections, the majority of participants attend grade 9 at the high school ( $n = 20$ ), with fewer attending grade 10 ( $n = 10$ ) and grade 11 ( $n = 6$ ). The study sample predominantly comprises male participants ( $n = 22$ ) compared to female participants ( $n = 14$ ), all of whom immigrated to the US from Latin American countries ( $N = 36$ ) and have high oral proficiency in Spanish, except for one student ( $n = 1$ ) whose L1 is a spoken-only Guatemalan dialect significantly different from Spanish.

### Data Collection

The study was conducted for the duration of an academic quarter in Fall 2023, with data collection period from October 26th, 2023 to December 15th, 2023. Institutional review board (IRB) approval for the study is attached in the Appendix A. Study timeline is presented in Table 2.

**Table 2:** Data Collection Timeline

Timeline	Type of data	Collected Data
October 26th- Nov 2nd <b>Training</b>	Pre-test	
		1. Background information
		2. WTC questionnaire
November 2nd-November 20th <b>AI desktop</b>	During the activity	
		3. Recordings of each AI interaction on desktop
Nov 20th <b>AI desktop post test + The same day VR Training</b>	Post test (after AI desktop)	
		4. WTC questionnaire
November 20th- December 18th <b>AI in VR</b>	During the activity	
		5. Recordings of each AI interaction in VR
December 15th, <b>Ai in VR post test</b>	Post test (after VR)	
		6. WTC questionnaire

In the pre-test phase, participants completed two online *Qualtrics* questionnaires distributed by their teachers in *Google Classroom*: first, Language and Social Background Questionnaire adapted from Anderson (2016) (Appendix B), Willingness to Communicate (WTC) Questionnaire adapted from Dewaele and Dewaele (2017) (Appendix C). These questionnaires were previously validated and used in numerous studies on willingness to communicate. All questionnaires were translated in Spanish, responses to the questions in the questionnaires were expected in Spanish as well. For students with lower literacy in Spanish, English version was provided with answers expected in English.

The intervention phase was divided into two sub-phases (desktop vs. VR), each beginning with training and ending with a post-test. An essential step in the study is teacher training using *Immerse*, conducted on October 15th. Students underwent training twice during

the study. On the first day of data collection, participants were trained to use the *Immerse* desktop version to communicate with AI bots using their *Chromebooks*. The training included accessing the application in *Google Classroom*, allowing access to their microphone in the *Google Chrome* browser, and navigating the home screen and selecting AI bots according to their level, topic, and preferred environment.

While teacher training that consisted of a two 3-hour-long session, was held before the beginning of the study, students were trained to use VR only before Phase 4. The training session on using VR focused on the basics of navigating in virtual reality using the *Meta Quest 2* headset and going through the *Immerse* tutorial to ensure that a) participants could assist the students in navigating the application effectively; b) data on students' interaction is not affected by the technology novelty factor or technical issues. Student training is presented in Figure 6.

**Figure 6:** Student Training using *Meta Quest 2* VR headsets



Although initially planned to include 6 interventions in total—three on desktop and three in VR—the final intervention phase was cancelled by teachers due to increased workload at the end of the semester, resulting in 5 interventions: three on desktop and two in VR. The study included two post-tests conducted after interacting with AI bots in different immersion types: Phase 2 for desktop and Phase 4 for VR. Each post-test elicited a qualitative dataset comprising video and audio recordings of each student’s 20-minute interaction with AI bots in low- and high-immersion environments.

For the low-immersion environment, participants recorded their screen, system audio, microphone, and front laptop camera using the embedded *Chromebooks* screen recording function. The recordings were automatically saved to a secured *Google Drive* folder with obfuscated video names, which was shared for the data analysis by the school IT department. In the high-immersion environment, participants recorded their screen, system audio, and microphone using the embedded recording function in the *Meta Quest 2* headsets. An Excel spreadsheet was used to track which VR headset was assigned to each student. At the end of the study, recordings were manually transferred to the researcher’s computer using a USB cable and assigned to students according to the information in the Excel file.

The post-test was conducted twice: first, after the intervention phase of conversational AI in the low-immersion environment; second, after the intervention phase in the high-immersion environment. Both times, data on participants’ willingness to communicate was collected.

### **Data Analysis**

After data was collected, it was analyzed to answer three research questions identified in this study.

### *RQ1: Impact of AI on WTC*

Research Question 1 (RQ1) aims to investigate the impact of conversational AI on willingness to communicate (WTC), based on the type of immersive environment (desktop vs. VR) and participants' ESL level. WTC is assessed through a WTC Questionnaire (Appendix C), administered to participants three times during the study: once as a pre-test to establish baseline, in Phase 3 as a desktop post-test, and in Phase 5 as a VR post-test. During the data analysis of the WTC Questionnaire, the percentage change in WTC was calculated as the difference in average percentage change, according to Equation 1. Percentage Growth in WTC 1 was calculated from the pre-test to the post-test after three desktop sessions. Percentage Growth in WTC 2 was calculated from the pre-test to the post-test after five sessions total, including two additional sessions in VR to previous three desktop sessions.

#### **Equation 1:** Percentage change

$$\text{Percentage change in WTC} = \left( \frac{\text{Post test} - \text{Pretest}}{\text{Pretest}} \right) \times 100$$

To determine the impact of ESL level and level of immersion on WTC growth, a Spearman correlation test was performed in PyCharm using Python, with imported libraries for data analysis and visualization: Pandas, Seaborn, Matplotlib, and NumPy. The Python code for Spearman correlation test is attached in Appendix D. It includes the code for calculating Spearman coefficients between WTC pre-test, WTC post-test 1, and WTC post-test 2, as well as correlation between percentage growth in WTC, ESL level, and level of immersion. Also, the code for visualizing the correlation in the form of a bubble graph is presented in the appendix.

## *RQ2: Language Production*

Research Question 2 (RQ2) investigates the effect of conversational AI on language production, considering the type of immersive environment, participants' ESL level, communication modality, and WTC.

Language production was captured from video recordings of students' interactions with AI bots. Each participant contributed a total of 5 complete data sets of language production samples ( $N = 5$ ), comprising three video recordings of interactions on desktop ( $n = 3$ ) and two in VR ( $n = 2$ ). Due to some students missing classes, actual data sets varied between three to five recordings per student. The duration of student interactions with AI bots ranged from one minute to twenty minutes based on proficiency level and their engagement in the conversation, with analysis focused on one conversation per session. Conversations were selected based on criteria such as completion of tasks or reaching maximum responses (10) per conversation. If multiple dialogues were recorded, the second dialogue was chosen to minimize novelty effects and technical issues typical for the first sessions.

Participants selected from 90 conversations spanning beginner to advanced levels across 9 immersive environments. Neither the environment nor the level of the conversation was predetermined and depended on teachers and students' preference. Despite instructions to speak to AI bots, 29 % of students ( $N = 10$ ) were typing their responses on desktop even after repeated instructions from teachers to speak.

To analyze the impact of participants' ESL level, preferred modality, WTC, and immersion level on language production, average percentage change in second language production was calculated. This metric considered the percentage difference in language production elicited from the five video recordings of each student's interactions with AI. All

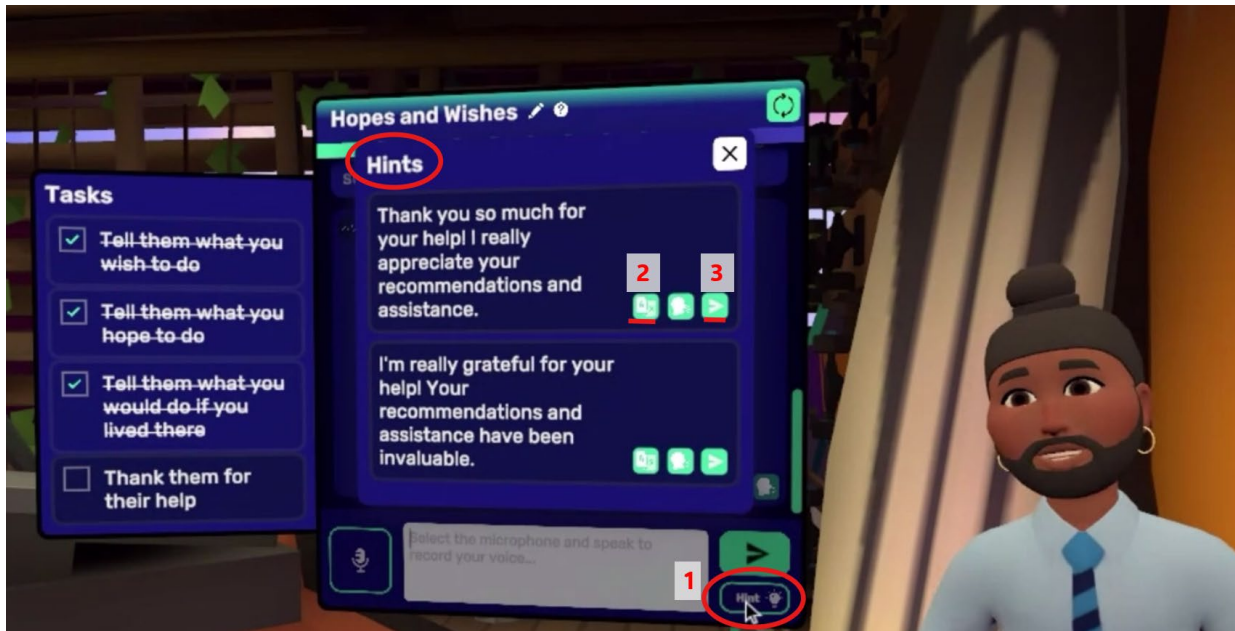
percentage changes in language production were averaged per student to determine their overall score of percentage growth in second language production. The score was used to calculate a Spearman correlation matrix to determine correlations between language production and listed variables. The Python code for calculating correlations between percentage change in language production with ESL level, WTC, level of immersion, and modality, is presented in Appendix D together with the code for visualizing the correlation matrix.

### *RQ2.1: Features of Conversational AI and Language Production*

To address RQ2.1 regarding the impact of AI conversational features used during interactions with the bots, eight features were identified from video recordings of student interactions. Six of these features were relevant for both levels of immersion, while two features were specific to the desktop version. The examined features are the following: hint, translate, send hint, define word, listen to the phrase, listen to the word, correcting writing, and Google Translate.

The hint button serves to help complete a specific communication task that is listed in the dialogue. Figure 7 illustrates the usage of the hint button by an ESL 1 participant. As they have the last incomplete task in the left tasks table “Thank them for their help”, the participant clicked on the hint button and was presented with two potential responses.

**Figure 7:** Conversational feature "Hints"

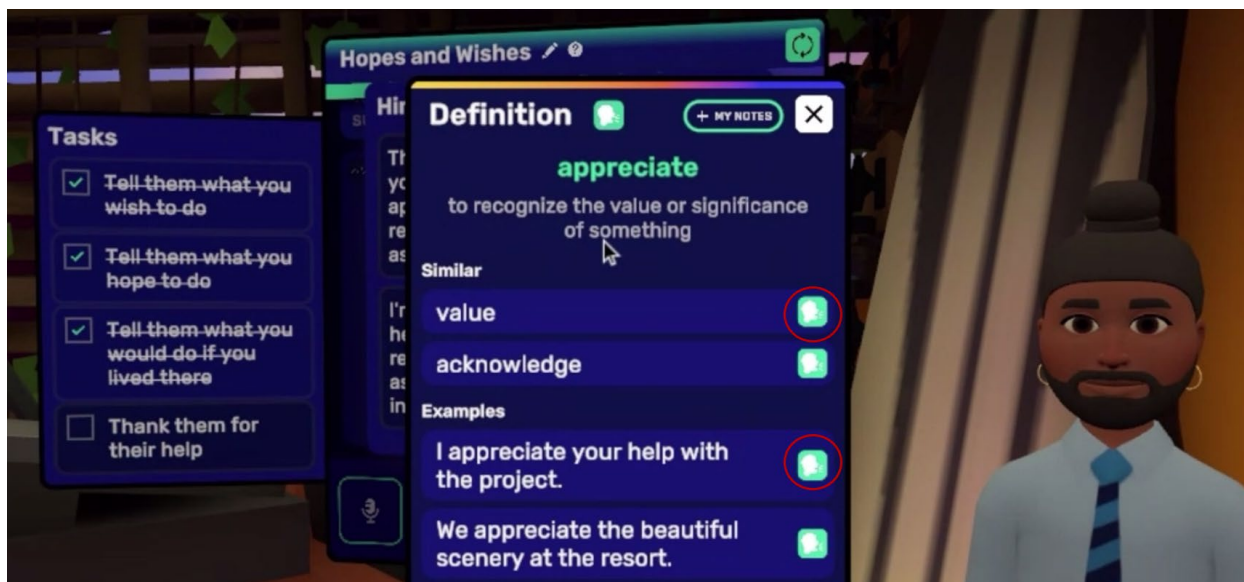


Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig4>

Number two in Figure 7 represents the translation feature generated to the participants' native language. Number three is the "send hint" function, which automatically sends a suggested phrase to the AI bot, eliminating the need to say or type your answer.

Other features illustrated by Figure 8 become available once you click on a word in the text. For instance, the participant clicked on the word "appreciate" and could see its definition, synonyms, and how this word is used in a sentence. Additionally, there is a button of a talking head which allows you to listen to the word or the whole phrase.

**Figure 8:** Conversational features "Word Definition" and "Pronunciation"



Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig4>

Finally, corrective writing and Google Translate are not embedded features but are considered as conversational tools that are actively used during conversations with AI. Corrective writing occurs when there is a discrepancy between the intended message and the perceived message by the AI, which could arise due to technical issues, poor sound quality, mispronunciation, or the presence of homophones in the target message. Despite the embedded translation feature in the application, some lower-level students still used Google Translate to construct their intended message in English. These two features are available only on desktop, as typing is not possible in the VR version of *Immerse*.

To determine correlations between language production and all features of conversational AI, the Spearman correlation test was conducted. This analysis included multiple variables such as language production, ESL level, modality, level of immersion, and all named conversational features. Similarly to the correlation test conducted for RQ1, the Spearman correlation test was

performed in PyCharm using Python programming language. The Python code for the Spearman correlation test and visual representation of the Spearman correlation matrix is attached in Appendix D.

Opposed to RQ2, language production in this analysis was not measured as average percentage growth. Instead, it was operationalized as the total number of syllables used to complete each conversation with AI. This approach aims to provide a better understanding of how conversational features impact language production in each session.

#### *RQ2.2: Features of Conversational AI and ESL Level*

To address RQ 2.2 regarding the conversational features used by lower-level ESL students, the Spearman correlation matrix was utilized using the same data. Specifically, the matrix incorporates variables for ESL level and the following features: hint, translate, send hint, define word, listen to the phrase, listen to the word, correcting writing, and Google Translate. The Spearman correlation coefficients were calculated to determine the correlation between ESL level and each feature. Additionally, visual representation was provided for data with a p-value lower than 0.05.

#### *RQ3: Recurring patterns of communication disruption with AI*

RQ3 investigates recurring patterns of instances when participants' communication with AI bots was disrupted. The patterns were formulated based on qualitative data from video recordings and classroom observations. Two patterns analyzed in this research question are limited recognition of mispronounced words and limited recognition of contextual cues. Mispronunciation is defined in this study as the incorrect articulation or pronunciation of a word due to phoneme substitution or sound omission. Contextual cues are defined as visual and auditory signals that students receive in the immersive environment, including questions asked

by AI, the chosen scene where the conversation occurs, and tasks given to students. Illustrative examples of communication breakdowns captured on video recordings were qualitatively analyzed, referring to possible linguistic explanations for the communication breakdown with the AI bot.

## CHAPTER 4: FINDINGS

### Desktop preference over VR headsets

Even though the initial study design had an equal number of sessions for each type of immersive environment (three on desktop and three in VR), the final number of students using VR was significantly lower than desktop users for several reasons. Not only was the last VR session canceled due to instructors' busy schedule, but also the majority of participants preferred the desktop version over VR. While all students were trained to use the *Meta Quest 2* VR headset before the VR intervention, Table 3 shows that only 53 % of students agreed to use VR headsets once during the study, and 22 % of students used VR twice as intended.

**Table 3:** ESL Level and Students VR usage during the study

ESL Level	Number of Students	% of Ss using VR at least once	% of Ss using VR two times
ESL 1	9	22%	22%
ESL 2	9	44%	0%
ESL 3	8	63%	13%
ESL 4	10	80%	50%
<b>Total</b>	<b>36</b>	<b>53%</b>	<b>22%</b>

Out of 53 % of students using VR, the highest number of students preferring VR was 80 % of ESL 4 students, and the lowest number was from ESL 1 level. During the second time using VR, only the ESL 1 group maintained the percentage of students using it. ESL 2 dropped from 44 % of VR users to 0 %, and ESL 3 decreased from the majority of students using VR to only 13 %. The only group that still showed a preference for VR was ESL 4, with 50 % of students using VR two times.

### RQ1: Impact of AI on WTC

RQ1 aimed to examine the impact of conversational AI on students' willingness to communicate, based on their ESL level and their preferred level of immersion. This research

question differentiated between the roles ESL level and level of immersion played in WTC scores and average percentage growth in WTC after three sessions on desktop and two additional sessions on either desktop or in VR.

Table 4 shows participants’ changes in willingness to communicate in English based on their preferred level of immersion. Since all students used the desktop version to converse with AI during the first three sessions, differences between the pre-test and post-test 1 are presented solely for desktop. However, once VR was introduced for students to converse with AI in Virtual Reality during the last two sessions, the majority of participants did not use VR headsets at all after the training. Specifically, 18 out of 36 participants used desktop for the last two sessions ( $n = 18$ ), ten students used VR once out of the two intended sessions ( $n = 10$ ), and eight students used VR two times as initially intended ( $n = 8$ ).

**Table 4:** WTC growth based on level of immersion

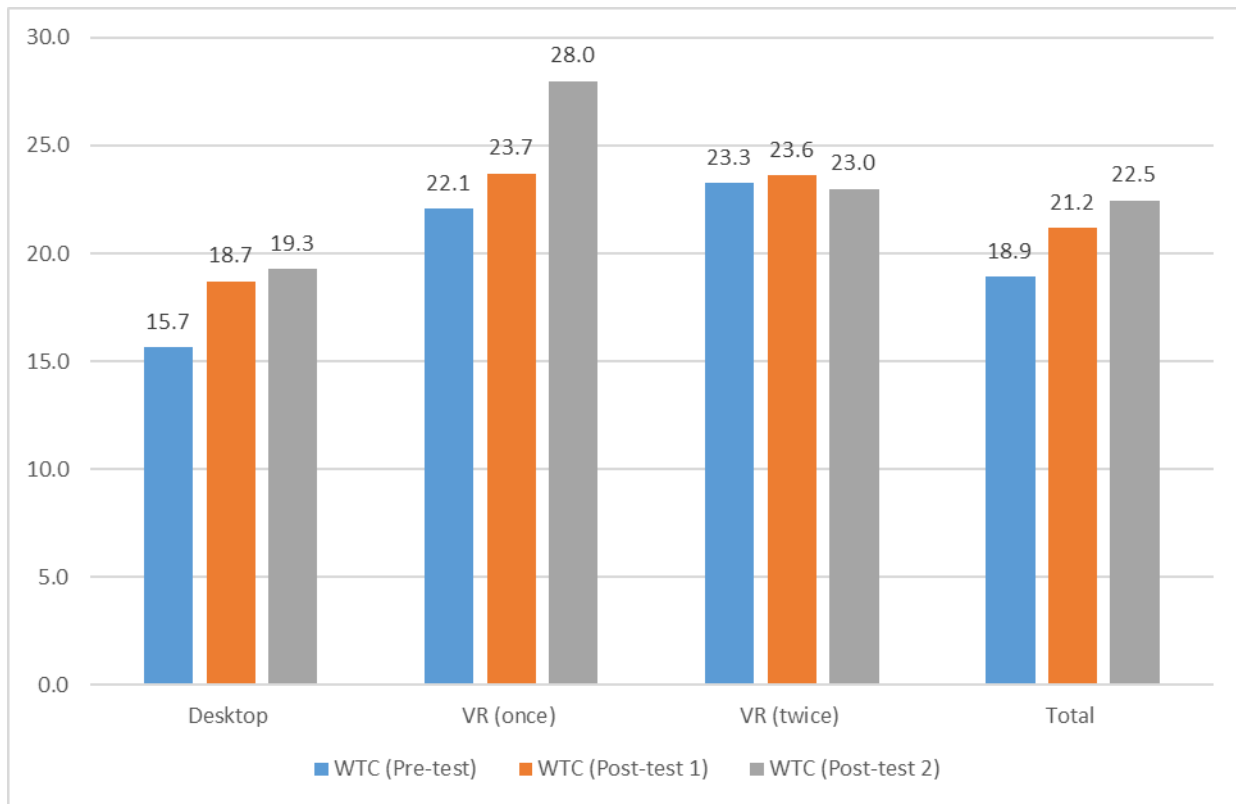
Number of sessions on desktop versus in VR				WTC (Pre-test) (n = 30)	WTC (Post-test 1) (n = 34)	% Growth WTC (Pre- Post test 1) (n = 28)	WTC (Post-test 2) (n = 11)	% Growth WTC (Pre- Post test 2) (n = 10)
Desktop 5 sessions VR 0 sessions “Desktop only” group (n = 18)				15.7	18.7	19%	19.3	23%
ESL 1 (n = 7)	ESL 2 (n = 6)	ESL 3 (n = 3)	ESL 4 (n = 2)					
Desktop 4 sessions VR 1 session “VR once” group (n = 10)				22.1	23.7	7%	28.0	27%
ESL 1 (n = 0)	ESL 2 (n = 3)	ESL 3 (n = 4)	ESL 4 (n = 3)					
Desktop 3 sessions VR 2 sessions “VR twice” group (n = 8)				23.3	23.6	2%	23.0	-1%
ESL 1 (n = 2)	ESL 2 (n = 0)	ESL 3 (n = 1)	ESL 4 (n = 5)					
<b>Total (N = 36)</b>				<b>18.9</b>	<b>21.2</b>	<b>12%</b>	<b>22.5</b>	<b>19%</b>

The highest percentage change in WTC from pre-test to post-test 1 is 19 % in the desktop group. The majority of students in this group are ESL 1 and ESL 2 level students. The lowest growth from pre-test to post-test 1 is in the “VR twice” group, with a 2 % growth in WTC. Based on all participants’ percentage growth, there is a 12 % growth in their willingness to communicate after three sessions conversing with AI bots on the *Immerse* platform.

For the percentage change in WTC from pre-test to post-test 2, conducted at the end of the study after 5 interventions, participants who used desktop all five times showed a 23 % growth in their willingness to communicate. Although the VR group using headsets once showed an even higher growth of 27 %, there is only one student who completed the post-test 2 WTC questionnaire from that group; hence, the percentage growth calculated for one student has limited applicability to the whole group. However, participants using VR two times, as initially intended for all participants in the study design, showed a 1 % decrease in their willingness to communicate.

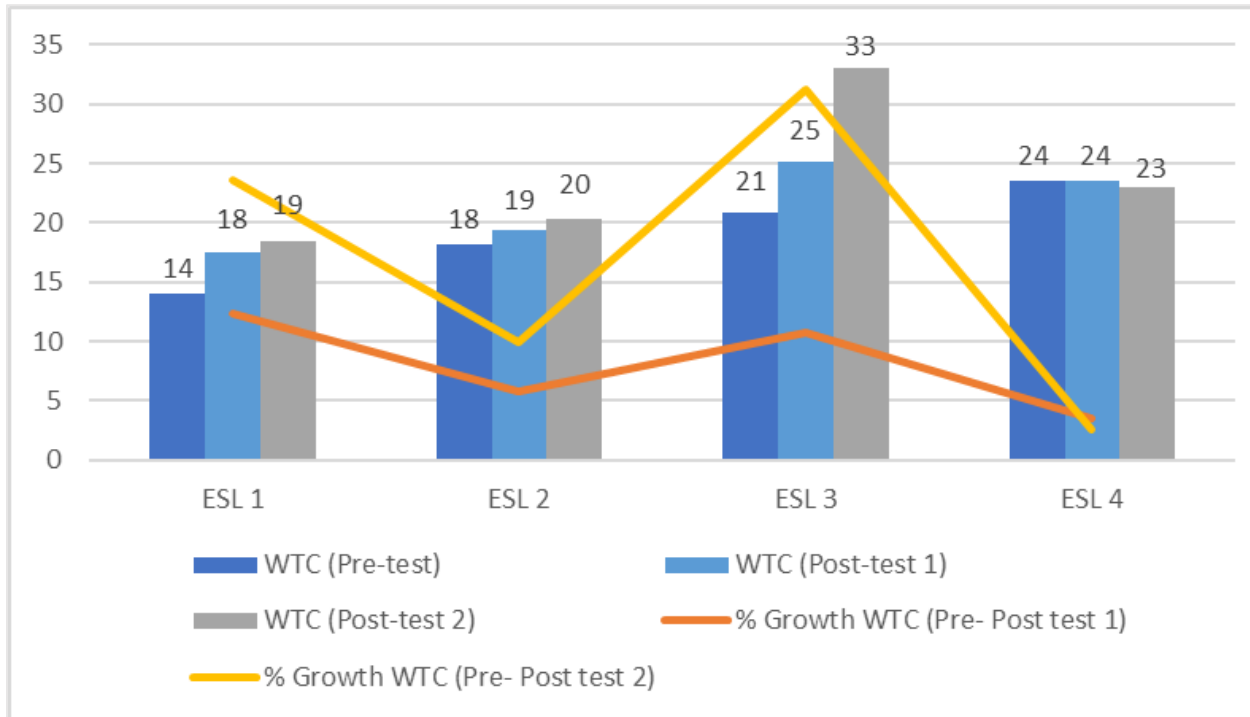
Figure 9 shows a change in WTC scores, with 30 being the highest possible score. The desktop group has the lowest pre-test score of 15.7, while the VR group using headsets twice received the highest score of 23.3. However, after five sessions, the VR group's score went down by 0.3% from the pre-test, while other groups showed a steady increase in WTC. Looking at all the participants regardless of their preferred level of immersion, their WTC increased by 12 % after three sessions of conversing with AI and by 19 % after five sessions, showing a positive impact of conversational AI on students’ willingness to communicate.

**Figure 9: WTC Scores for Desktop and VR Groups**



According to the data, there is an apparent preference for a low-immersion environment among lower-level ESL students, while advanced ESL students prefer the high-immersion VR. This preference can be inferred from the similarities observed in Figure 10, which illustrates the percentage growth in WTC based on participants' ESL level. In Figure 10, columns represent participants' WTC scores for the pre-test and two post-tests, while lines indicate the percentage change in their WTC scores from the pre-test to each post-test.

**Figure 10:** Percentage Growth in WTC based on ESL level



Comparing Figure 9 and Figure 10, the columns in Figure 10 representing scores for ESL 1 and ESL 2, show a similar pattern to the columns in Figure 9 representing the desktop group. Additionally, the ESL 3 group from Figure 10 exhibits a pattern similar to the “VR once” group from Figure 9, showing a significant increase in WTC scores from the pre-test to post-test 2. Finally, the ESL 4 group from Figure 10 demonstrates a decline in WTC scores similar to the “VR twice” group from Figure 9. This comparison underscores the significance of the Spearman correlation test, presented in Figure 11, which is important for determining whether it is the level of immersion that correlates with the students’ WTC, or if it is solely their ESL level.

**Figure 11:** Spearman Correlation WTC, ESL level, and level of immersion

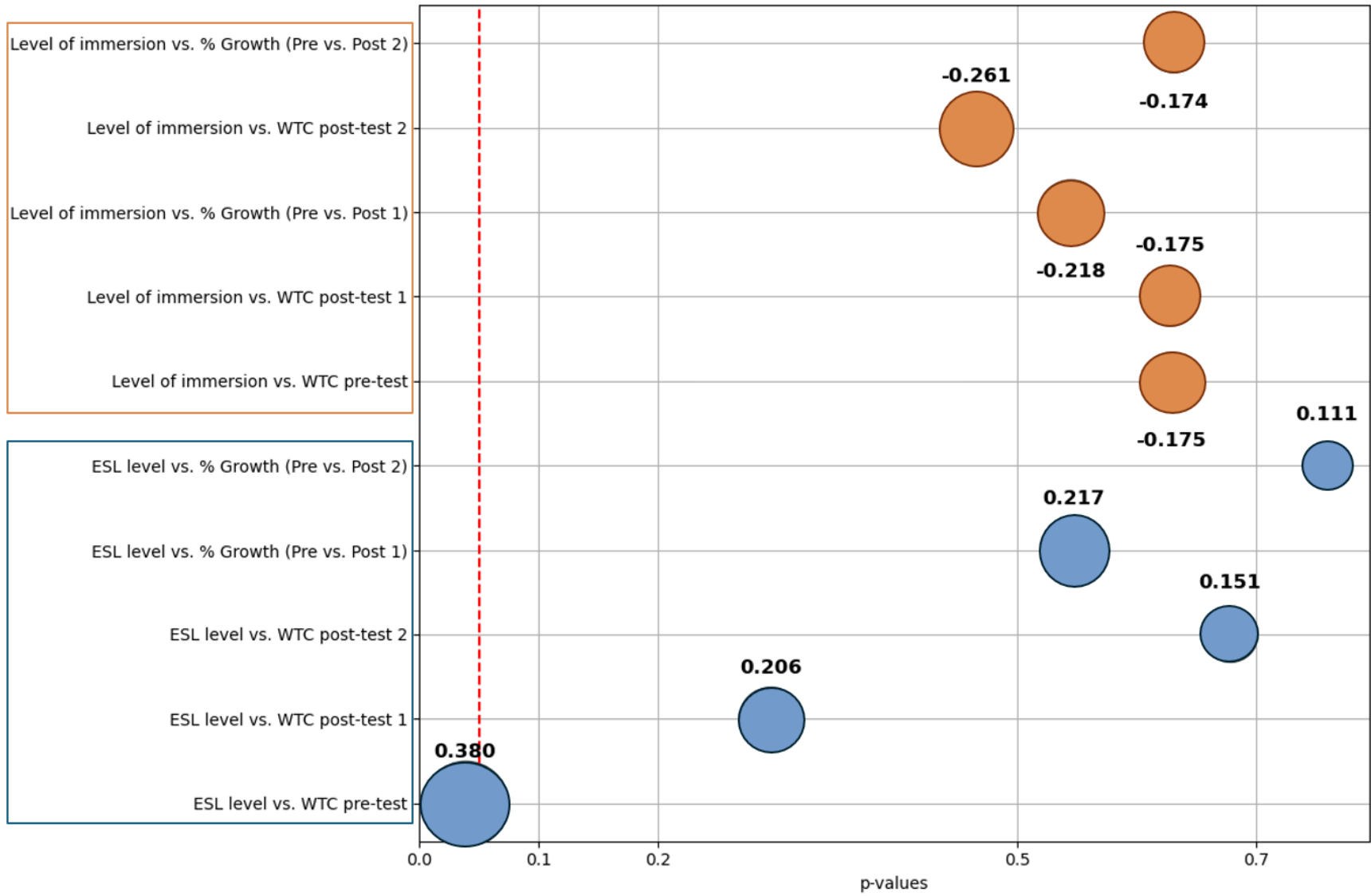


Figure 11 represents Spearman's correlation test for the level of immersion and ESL level with WTC test results. Blue bubbles represent correlations between ESL level and five previously analyzed WTC data points, including the pre-test, post-test 1, post-test 2, and percentage growth from pre-test to each post-test. Orange bubbles represent correlations between levels of immersion with the same WTC scores. Bubble size is determined by the Spearman correlation coefficient, the higher the correlation is, the bigger is the bubble. The blue color indicates a positive correlation, and the orange one indicates a negative correlation. The bubbles are positioned on the figure according to their p-values, with a red line representing the threshold for a p-value of 0.05.

According to the figure, the correlation between multiple WTC scores and participants' ESL level is positive, meaning that as participants' ESL level grows, their WTC scores also increase. However, the only statistically significant correlation is between ESL level and the WTC pre-test, with a Spearman correlation coefficient of 0.38 and a p-value of 0.038. While the Spearman correlation coefficient is relatively significant (above 0.2) for correlations between ESL level and WTC pre-test 1, and ESL level with WTC percentage change from pre-test to post-test, the p-values are over 0.2 and 0.5, respectively, indicating that these correlations are not statistically significant.

Regarding the level of immersion, there is no statistically significant correlation between WTC scores and the preferred level of immersion. Nevertheless, there is a trend indicating a negative correlation between WTC scores and the level of immersion, suggesting that participants with higher WTC scores used VR more than those with lower WTC scores. The lack of a statistically significant correlation, especially between these two variables and WTC post-

test 2, can be explained by gaps in data, as only 11 participants completed the final WTC post-test 2 compared to 34 participants who completed the WTC post-test 1 questionnaire.

In response to RQ1, conversational AI had an overall positive impact on students' willingness to communicate, resulting in a 12 % increase in WTC after three sessions and a 19 % increase after five sessions. Participants who used the desktop version for the entire study duration showed greater gains in WTC of 23 % compared to the group that used VR twice, which showed a 1 % decline in their WTC score. Moreover, lower-level ESL students preferred the desktop version compared to higher-level ESL students, who tended to choose VR more often. While lower-level students had lower WTC scores according to the Spearman correlation test, their WTC growth was more significant than that of higher-level students. Therefore, conversational AI positively affected participants' willingness to communicate, with more significant growth in WTC observed among ESL 1 and ESL 2 students.

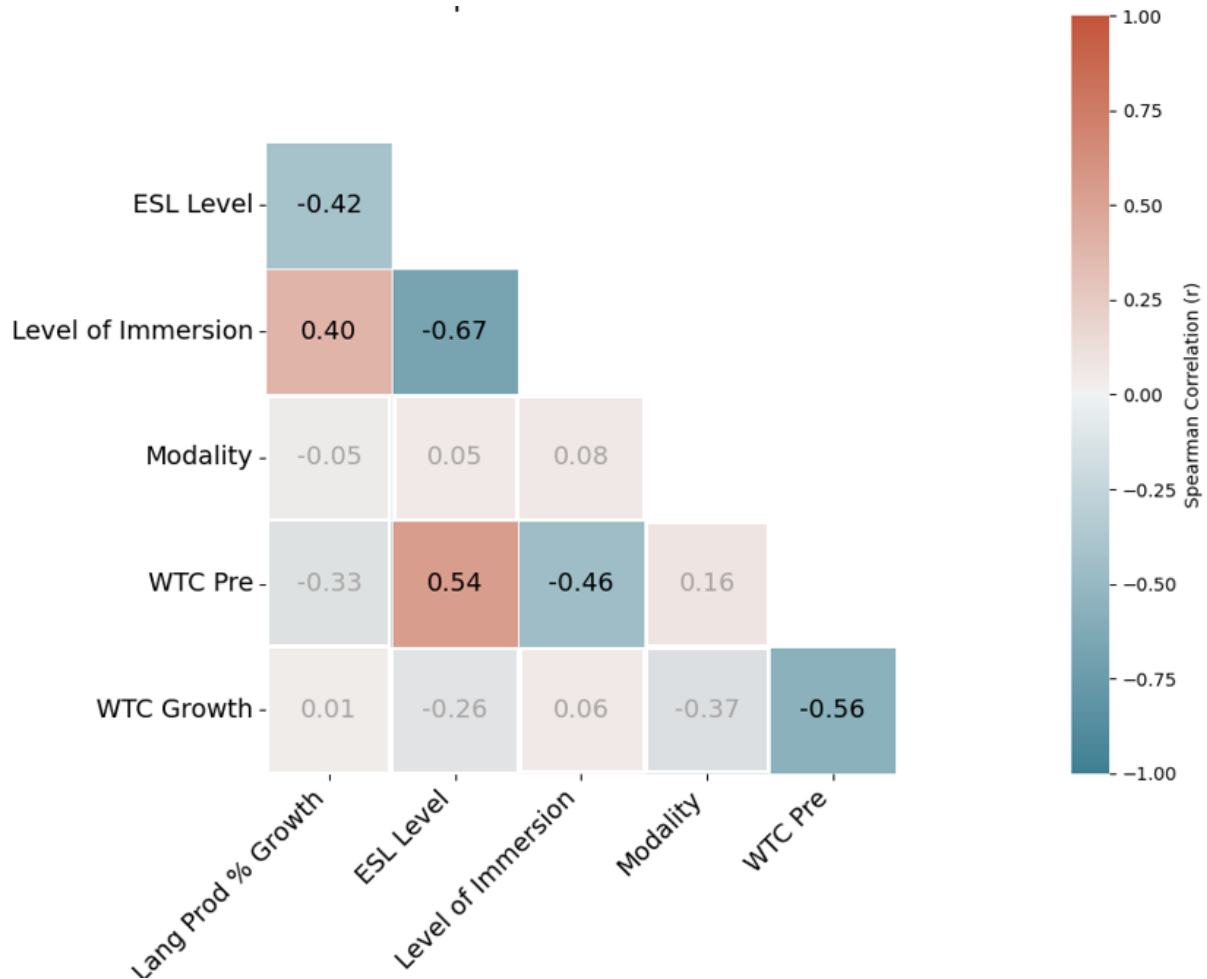
### **RQ2: Impact of AI on Language Production**

RQ2 aimed to investigate how language production changed during five sessions of conversations with AI bots. To explore the effect of conversational AI on the average percentage change in language production based on participants' ESL level, preferred communication modality, level of immersion, and willingness to communicate, a Spearman correlation test was conducted and the results are presented in Figure 12.

The average percentage growth in language production for all students was 49 %. Figure 12 displays Spearman correlation coefficients between the percentage growth in language production and ESL level, level of immersion, modality, WTC pre-test, and percentage growth from WTC pre-test to post-test 1. All data are color-coded, with the red color representing positive values and the blue color representing negative ones. The intensity of the color signifies

the strength of the Spearman coefficient between two variables, which is also indicated inside the colored squares. Boxes that have a grey color indicate p-values higher than 0.05, suggesting lower significance in the correlation between these variables.

**Figure 12:** Spearman correlation matrix for percentage growth in language production



To calculate the Spearman correlation, some variables were converted into numeric values. For the level of immersion, VR is represented by 1 and desktop by 2. Regarding modality, spoken modality is represented by 1 and written by 2.

The highest negative correlation found was between the average percentage growth in language production and participants' ESL level:  $r = -0.42$ ,  $p = 0.025$ . This indicates that lower-level participants experienced higher growth in their language production from the first session

to the next one till the last session. Additionally, there is a positive correlation of 0.4 with a p-value of 0.035 between the level of immersion and language production, meaning that participants using the desktop interface usually have higher average growth in the amount of language they produce. However, while the Spearman coefficient is relatively high for WTC pre-test, the p-value is 0.088, indicating that this relationship is not statistically significant.

According to the correlation matrix, a strong negative correlation of -0.67 was also found between ESL level and level of immersion, indicating that students with higher ESL levels tend to use VR more than the lower-level students. Moreover, WTC pre-test scores positively correlate with students' ESL level,  $r = 0.54$  with a p-value of 0.003, indicating that lower-level students tend to have lower initial willingness to communicate. Conversely, there is a negative correlation between WTC pre-test and level of immersion of -0.46 with a p-value of 0.013, showing that students who prefer VR over desktop also have higher initial willingness to communicate. However, percentage growth in willingness to communicate from pre-test to post-test 1 does not have a statistically significant correlation with other variables except for WTC pre-test scores.

Although no significant correlation was found between percentage growth in language production and modality, the average percentage growth in language production for students engaged in written communication is 14 %, while it is 59 % engaged in spoken communication.

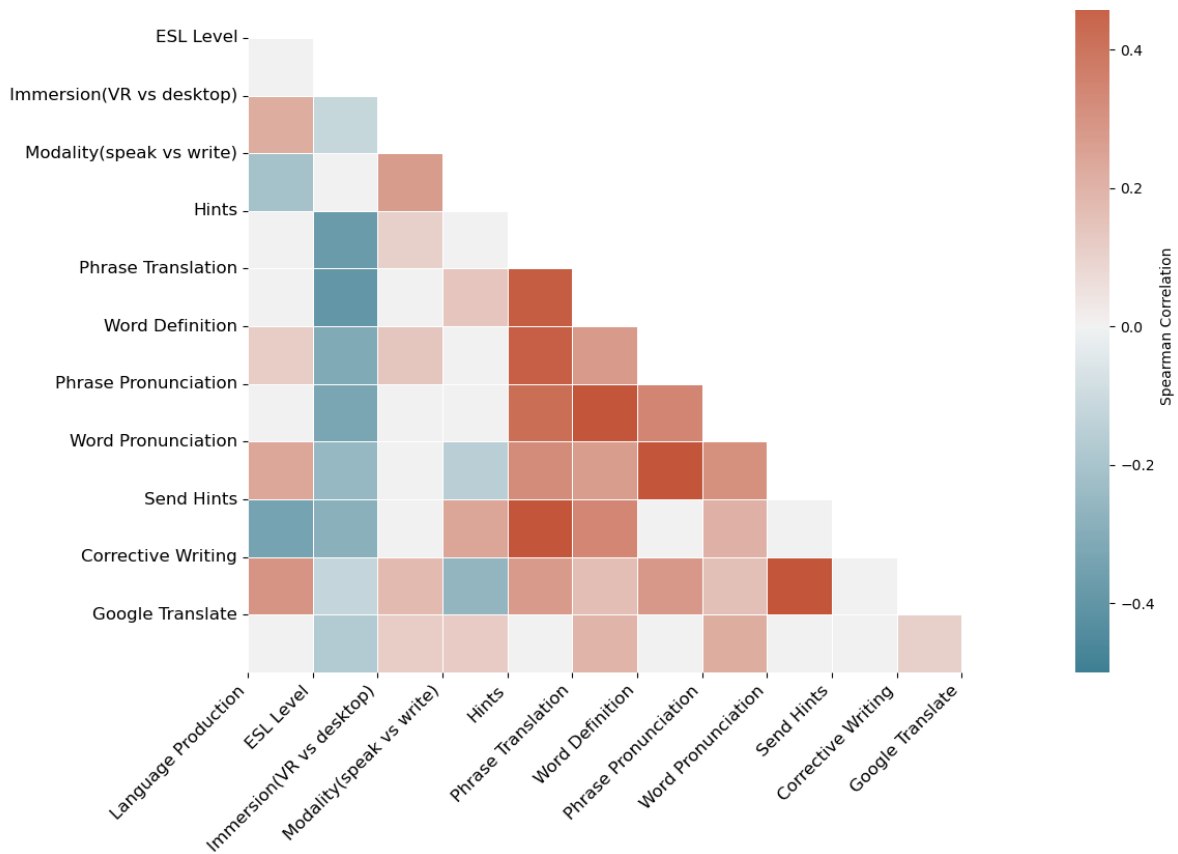
In summary, lower-level ESL students using the desktop version of the platform tend to have higher percentage growth in second language production. Other variables such as modality and WTC scores do not significantly impact the percentage growth in language production.

*RQ2.1: Features of Conversational AI and Language Production*

Research Question 2.1 (RQ2.1) aimed to explore the effect of six conversational features on language production in each conversation with an AI bot for each student across ESL levels and communication modalities.

Figure 13 displays the Spearman correlation matrix for conversational features and language production. The intensity of correlation is indicated by color: red denotes a positive correlation, grey signifies a low correlation, and blue represents a high negative correlation. This figure presents all correlations with high Spearman coefficients that have not been adjusted for p-values.

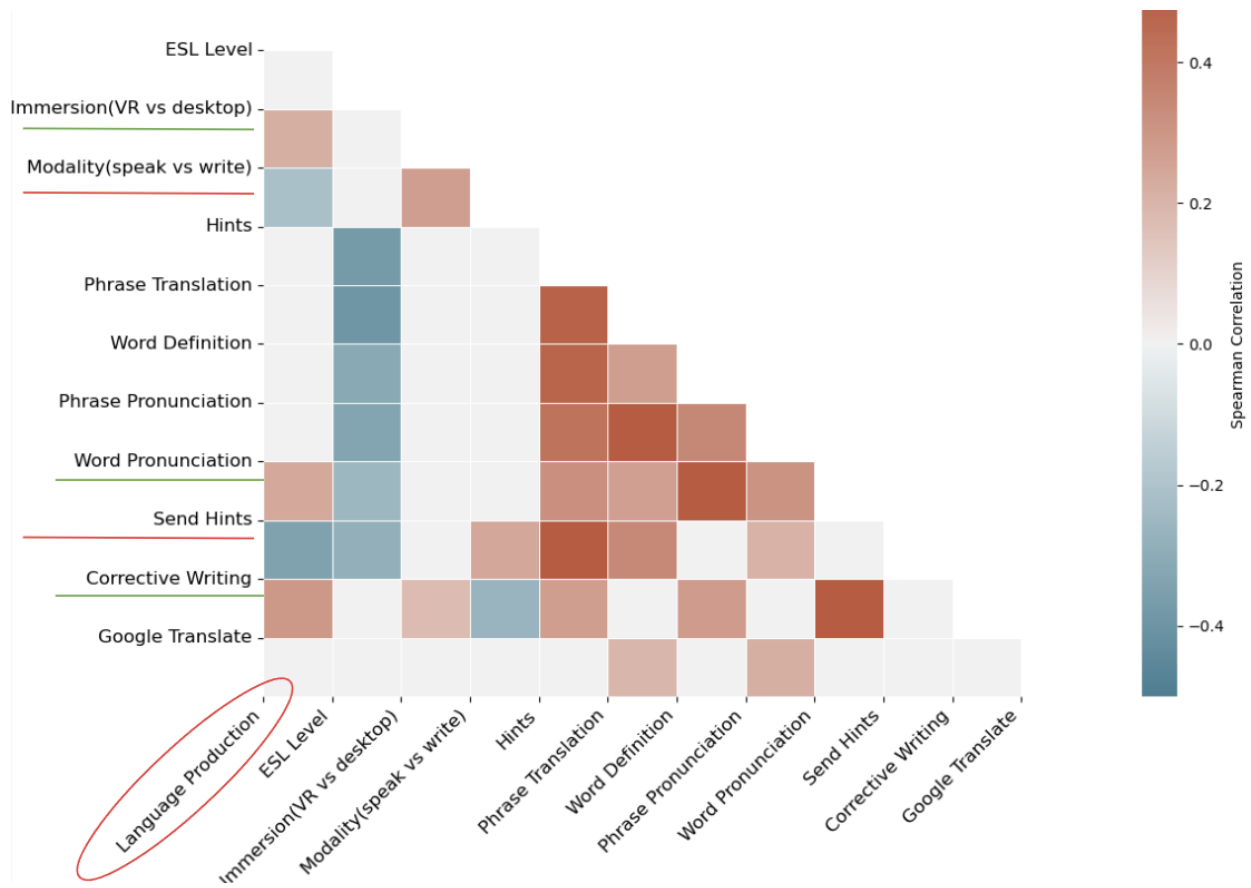
**Figure 13:** Spearman correlation matrix for conversational features and language production



However, Figure 14 is based on the same data but only represents correlations with p-value lower than 0.05. According to Figure 14, language production has a weak positive correlation with using a low-immersion environment ( $r = 0.22$ ,  $p = 0.009$ ), which was also observed for the percentage growth in language production. However, ESL level, which had one of the highest correlations with language production growth, shows a very low Spearman coefficient of  $-0.02$  and a p-value of  $0.82$ .

Conversely, for modality, while no correlation was found between modality and percentage growth in language production, there is a weak negative correlation of  $-0.2$  found between the written modality and the amount of language produced.

**Figure 14:** Spearman correlation matrix for language production and features with p-value  $< .05$



Regarding specific functions, there is a significant positive correlation between language production and listening to the pronunciation of a word ( $r = 0.24$ ), as well as using corrective writing when the intended message was not correctly interpreted by AI ( $r = 0.29$ ). However, the send hint function shows a negative correlation with language production, specifically  $r = -0.34$ . Interestingly, no correlation was found between using translation features and language production.

#### *RQ2.2: Features of Conversational AI and ESL level*

The same Spearman correlation matrix presented in RQ2.1, is used to answer RQ2.2 regarding the connection between students' ESL level and features used during each conversation with AI. According to the second column of Figure 14, which represents ESL level, there is a negative correlation with all the features. For instance, the strongest negative correlation,  $r = -0.39$  with a p-value of 0.000004, is observed with the phrase translation feature. The least significant correlation is with Google Translate, which has a negative correlation of -0.12; its p-value is on the threshold at  $p = 0.053$  but can still be considered statistically significant. Other coefficients range from -0.33 to -0.25, indicating that the lower the ESL level, the more students tend to use features that provide additional support to complete the conversation.

### **RQ3: Recurring patterns in human-AI communication breakdowns**

RQ3 investigates recurring patterns of instances when participants' communication with AI bots was disrupted. The patterns were formulated based on qualitative data from video recordings and classroom observations. Table 5 represents three recurring patterns.

**Table 5:** Recurring Patterns in communication disruption

<b>Recurring patterns</b>	<b>Example</b>	<b>When occurs</b>	<b>How often occurs</b>
<b>Limited recognition of mispronounced words</b>	For the word pants, students instead of [pænts] would say [pʌnts] while completing a role-play in the shopping center. AI recognized [pʌnts] as "bombs"	For lower-level ESL students, with specific words: pants, jeans, towel, about, family, cash	Out of all ESL 1 and ESL 2 levels who were primarily speaking with the bot, 78 % mispronounced a word at least once
<b>Limited recognition of contextual cues</b>	AI asked: "What are you doing tonight?" The student intended to say: "I'm meeting a friend tonight". Instead, AI recognized it as "I'm eating a friend tonight".	With homophones or near-homophones for primarily lower-level ESL students.	It happened at least once to 80 % of students who were speaking.

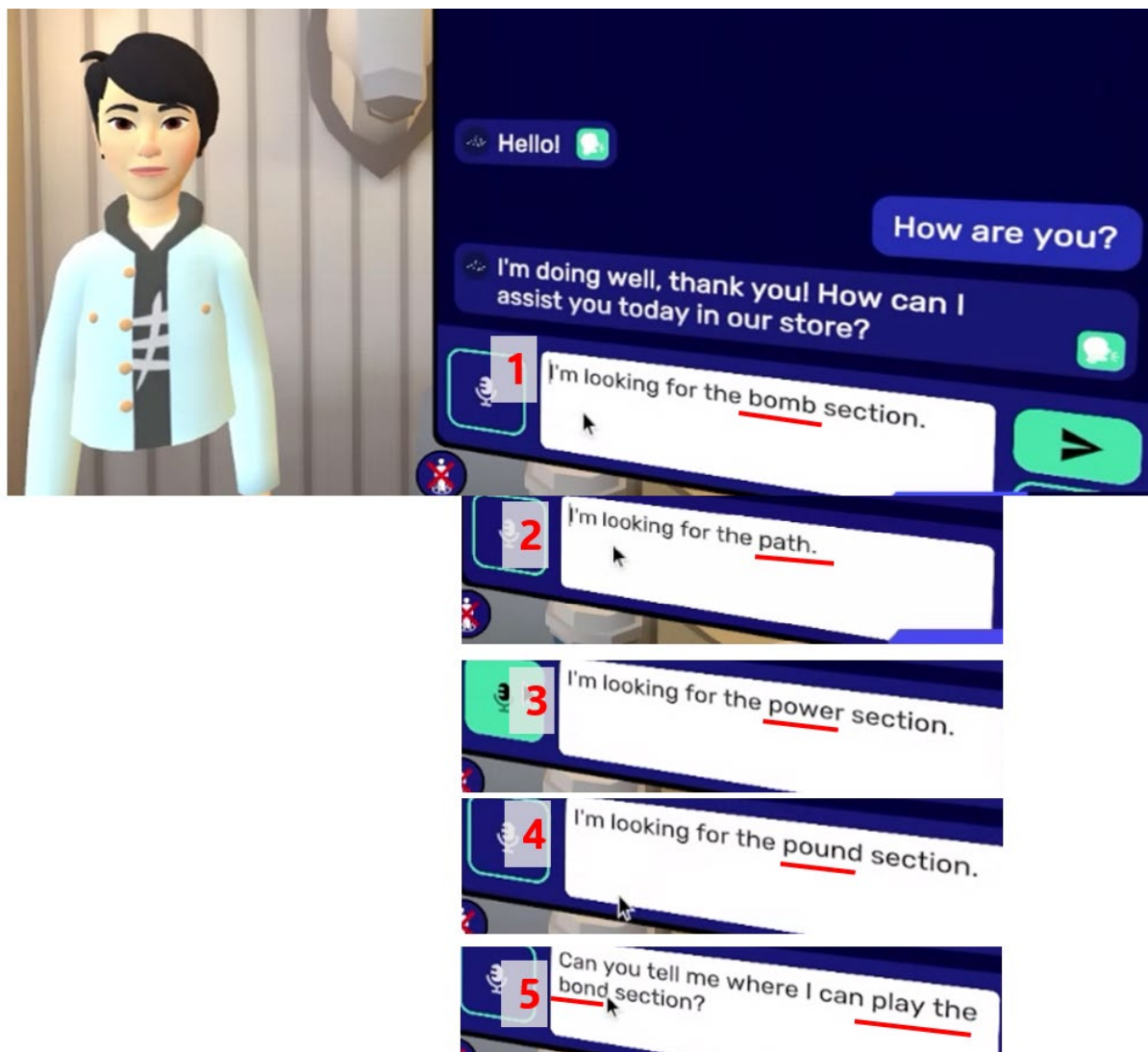
*Limited recognition of mispronounced words*

Natural Language Recognition (NLR) of the conversational AI used in this study accurately recognized accented speech. However, when it came to mispronounced words, lower-level students struggled the most in conveying their intended messages. While students typically repeated the same phrase to help the AI understand them, some also corrected perceived errors by typing the correct version, despite primarily communicating orally. This especially occurred if the AI failed to grasp their intended message after two attempts. One word that proved particularly challenging for many Spanish-speaking students, and which was frequently repeated during the study, was "pants", referring to a role-play scenario in a shopping center.

Figure 15 presents an illustrative example of limited recognition of mispronounced words, showing a conversation of an ESL 2 student and an AI bot in a shopping center on a beginner-level topic of "Costs and Purchases". Numbers one to five in the figure show students' responses to the conversational AI to complete the task of asking where the pants section in the shopping center is located. The student had several pronunciation issues with the word "pants"

that are typical for Spanish-speaking ESL students. First of all, the word "pants" contains a cluster of three consonants at the end, which is not typical in Spanish and makes it challenging to pronounce. Moreover, Spanish speakers, especially in the Caribbean and some parts of Latin America, have an aspirated [s] sound at the end of words or may drop it, which can make it difficult to articulate [pænts]. Finally, the [æ] sound in "pants" is not typical for Spanish and was substituted by the student with [ɑ], resulting in a different pronunciation of the word.

**Figure 15:** Limited recognition of mispronounced words



Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig13>

According to the video recording, represented in the Figure 15, the student attempted to say, "I am looking for the pants section." However, the AI recognized the word "pants" as "bomb," "path," "power," "pound," and "bond." On the fifth attempt, the student intended to say, "Can you tell me where I can find the pants section?" but the AI didn't recognize not just the word "pants" but also the word "find." After the fifth attempt, the student changed the word "pants" to "jeans" and sent the message, where "jeans" was perceived as "children's." Despite producing more language to complete the conversation through repetition, the student still could not convey the intended message. Even though the student understood that the issue was with the word "pants," she did not listen to the pronunciation of the word "pants" or to the whole phrase suggested in the hints.

The limited recognition of mispronounced words is further exacerbated by the second observed pattern regarding contextual cues.

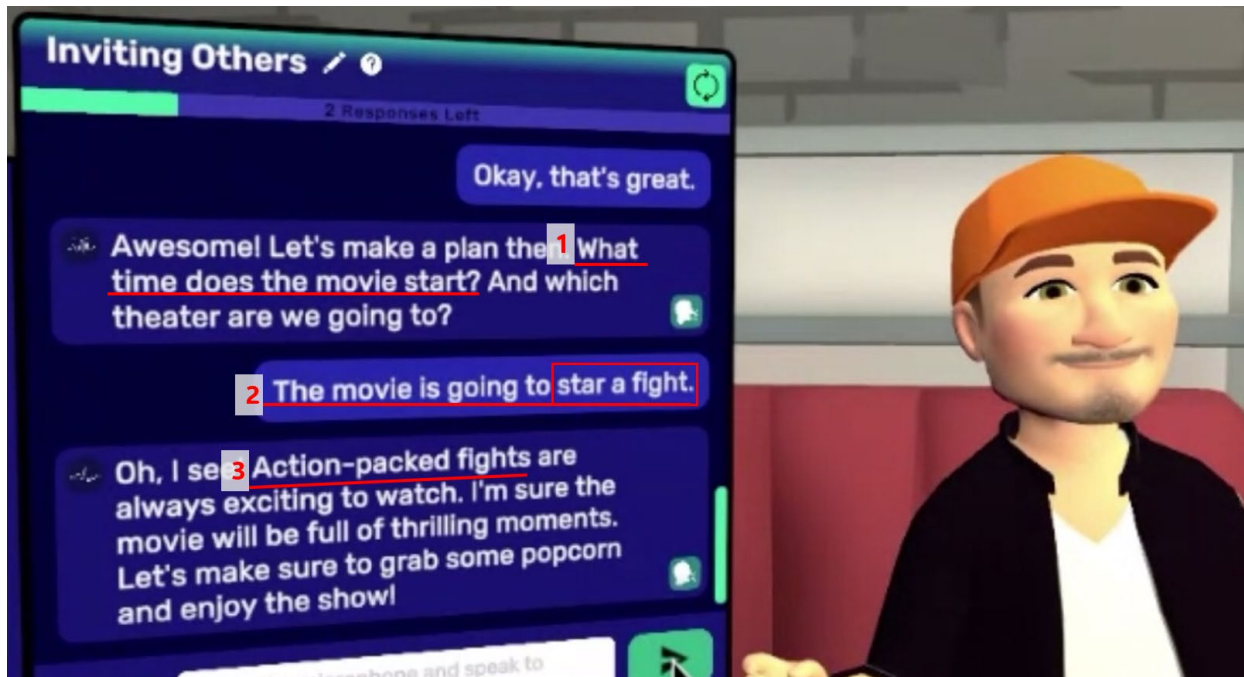
#### *Limited recognition of contextual cues*

Human conversation partners rely on a number of contextual cues during conversations to decode the intended message. Contextual cues provide insight into the current situation. During conversations, people do not solely rely on what they hear but also on visual and auditory contextual cues. However, the conversational AI used in this study has a limited ability to recognize and interpret contextual cues, which led to communication breakdowns.

In Figure 16, we see a dialogue of an ESL 2 student with the conversational AI. Students' responses are on the right side of the dialogue window. The conversation, on an intermediate-level topic of "Inviting Others," takes place in a fast-food restaurant. Number 1 in the figure indicates a question asked by the AI bot, "What time does the movie start?" The student recorded their response by saying, "The movie is going to sta a fi," intending to say that the movie is

going to start at five. The "sta a fi" part was recognized by Natural Language Recognition (NLR) as "star a fight" instead of "start at five," which can be seen under number 2 in the figure. The limited recognition of contextual cues resulted in a shift in the topic response, where the AI generated a sentence about "action-packed fights."

**Figure 16:** Limited recognition of contextual cues, ESL 2 example

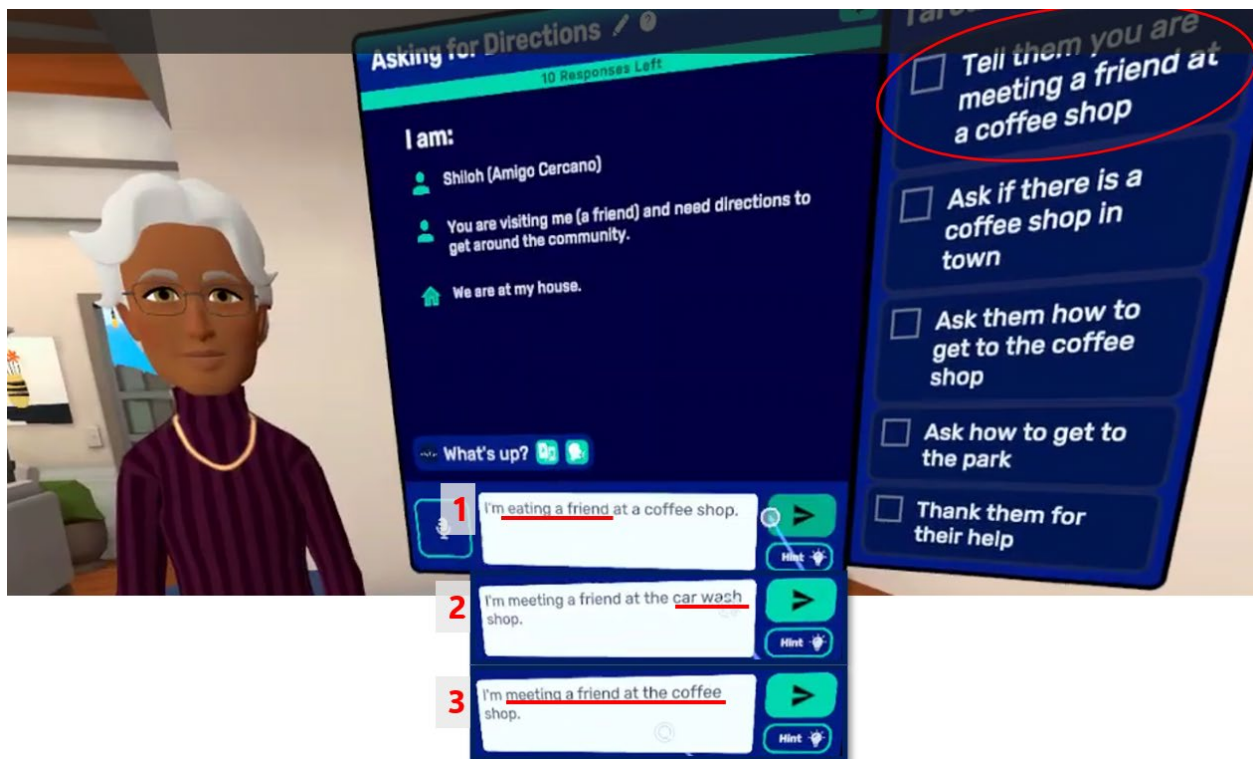


Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig14>

Another illustrative example is from an advanced-level dialogue in the home scene involving an ESL 4 student who did not have pronunciation issues. Figure 17 shows all responses of the student as they completed the first task for the conversation: "Tell them you are meeting a friend at a coffee shop." Throughout the conversation, the student's intended message remained consistent: "I'm meeting a friend at the coffee shop." However, the phrases "meeting a friend" and "eating a friend" exhibit phonetic similarity or near homophony. The student utilized connected speech, specifically gemination, where the [m] consonant sound extended across a

word boundary, resulting in a lengthened [m] sound in "I'm meeting" instead of two distinct [m] sounds. While connected speech is indicative of fluency in ESL, natural language recognition did not perceive the intended message due to gemination, and the AI did not consider contextual cues such as the task the student had or the unrealistic nature of "eating a friend," which is highly unlikely in conversation.

**Figure 17:** Limited recognition of contextual cues, ESL 4 example



Note: The video version can be accessed through the link: <https://tinyurl.com/uo2fig15>

For the second attempt, the student emphasized the misinterpreted word "meeting," making two distinct [m] sounds, resulting in the recognition of the phrase "I'm meeting". However, while focusing on the word "meeting", the student said the word "coffee" faster than before, blending it even more with the word "shop", which resulted in a perceived response of "car wash shop" instead of "coffee shop". This could again be caused by connected speech. For

example, vowel reduction in the word "coffee" could result in a change from [ɔ:] sound to [ə] sound, resulting in a pronunciation of the word "coffee" as ['kəfɪ]. Additionally, the [f] sound in "coffee" might have been elided or dropped. During the final attempt, the AI fully recognized the intended message. However, the time that it took the student to produce each phrase increased from 205 hundredths of a second to 221 hundredths of a second on the second attempt, and to 248 hundredths of a second on the final attempt. Not only did the student slow down, but he also limited his usage of connected speech, particularly consonant blending.

Summarizing these two patterns of limited recognition of mispronounced words and contextual cues, it was not always clear if the intended message was not recognized by the conversational AI because of mispronunciation of the word or because of the lack of recognition of contextual cues—it is often both, especially for lower-level students. Nevertheless, the pattern of AI not recognizing the intended messages that could be understood from the context by a human communication partner happened at least once to 80 % of students who chose to communicate with AI orally. While recognition of contextual cues occurred regardless of students' ESL level, the issue of recognizing mispronounced words was most apparent among low-level ESL students. Out of all ESL 1 and ESL 2 students who primarily spoke with the AI bot, 78 % faced issues with recognition of words that they mispronounced.

Nevertheless, there are both advantages and disadvantages to these patterns for language production based on participants' ESL level. For lower levels, both limited recognition of mispronounced words and contextual cues might raise their awareness of their pronunciation issues, as their speech is recognized in isolation from the environment and context. It can encourage students to articulate more clearly, confirm the correct pronunciation of a word they struggle with, and produce the intended message several times, thereby increasing their language

production. On the other hand, it can affect their confidence as English speakers and lead to frustration. However, considering quantitative findings, lower-level ESL students had both a higher positive percentage growth in language production and a higher positive percentage growth in willingness to communicate after interactions with AI. Moreover, the word pronunciation feature was the only one that had a high correlation with language production, showing a connection between students checking word pronunciation and the amount of language they produce.

Compared to some positive aspects of limited recognition of contextual cues for low-level ESL students, advanced ESL students seemed to be frustrated when AI did not recognize their intended message. They spoke less fluently and naturally than they usually do. Quantitative findings of a slightly negative percentage change in WTC among ESL 4 level students, as well as the found negative correlation between higher ESL level and language production (-0.42), can be partially explained by qualitative data on breakdowns in communication.

## CHAPTER 5: DISCUSSION

This study contributes to limited research on utilizing immersive conversational AI with the under-researched immigrant high-school population to improve their language production and willingness to communicate in ESL. The development of these two variables through this novel technology was observed through the ecological perspective and Complex Dynamic System Theory (CDST), emphasizing that both language production and willingness to communicate are not fixed for all learners but dynamically evolve and differ between individuals. Among a variety of internal and external factors that influence language production and WTC, this study identified that they varied based on factors students could control: preferred modality, level of immersion, usage of conversational features, difficulty level of the conversation topic, immersive environment, and their involvement in the activity. However, significant impacts also played factors outside of students' control: their teachers' perception of VR, technical issues with sound quality and speech recognition, as well as less tangible and obvious factors such as peer pressures and classroom layout.

While the findings in this study show a correlation between lower ESL levels and higher growth in language production and WTC, it is crucial to explore the complex interconnection of students, teachers, tools, and environment to have a full picture of how this new technology is used in the naturalistic environment by a vulnerable group of immigrant students, which is often overlooked in school-technology research (Carhill-Poza & Williams, 2020). The findings of the current study show that the growth in WTC correlates with lower ESL levels, usage of low-immersion VR, and initially lower WTC scores (RQ1). Average percentage growth in language production solely correlates with lower ESL levels and usage of desktop, but when we look at language production as the total number of syllables per conversation, it also positively

correlates with the spoken communication modality and listening to the pronunciation of individual words (RQ2, RQ2.1). However, usage of all features, including the ones that positively affect language production, highly correlates with the lower ESL level (RQ2.2). Finally, qualitative data show individual differences in observed variables. The same pattern in communication disruption usually occurs in lower-level students due to mispronunciation of some words, raising their awareness of word pronunciation and leading to higher language production through repetition, which also encourages the memorization of new words. However, advanced students usually encounter the same issue due to instances of fluent connected speech, where some sounds are blended. In this case, communication disruption leads to higher frustration and less fluent speech (RQ 3).

### **Preferred level of immersion**

Even though the initial study design intended to have an equal distribution between desktop sessions and VR sessions, teachers cancelled the last day of the study, which was planned for VR intervention. However, students' preference for desktop over VR was apparent: only 53 % of students used VR after the training, and only 22 % chose VR during the last two interventions. As students were asked to use only VR during the VR intervention phase, the desktop served as an alternative option if students could not use VR for health-related reasons. However, positive perception towards high-immersion VR, found in Huang et al. (2021), Thrasher et al. (2024), and Xie et al. (2019), was not as apparent in this study due to a strong preference for low-immersion VR by the majority of students.

Among the reported reasons for not using VR was primarily physiological discomfort, including motion sickness and headaches, which correlates with previous CALL research findings (Kaplan-Rakowski & Gruber, 2023; Thrasher, 2022; Thrasher et al., 2024). However,

female participants also frequently indicated that they did not want to use VR because of their makeup or hairstyle, which was not emphasized in previous VR studies. Also, female participants were particularly hesitant to use VR: only 2 out of 14 female participants used VR headsets twice during the study.

Another observed reason, which was not reported by students, relates to peer pressure in high schools. Peer pressure as a “prominent attribute of adolescence” (Clasen & Brown, 1985, p. 452) has been observed as a multidimensional concept affecting teenagers’ mental health and socialization process. This sort of high-school peer pressure also played a role in this study. Multiple instances against using VR in the study can be illustrated by an example of a female student from the first group, which had a total of six participants. The student was whispering in every desktop session, which correlates with the teachers’ description of the student as very “shy to speak”. After using VR once, the student was speaking loudly for the first time during the study and said that she felt less nervous, as she “could not see other students in the classroom”. However, the second time the students needed to use VR, other students did not want to use the headsets. Even though this particular student wanted to use VR, she hesitated and finally decided to use the desktop version as the rest of the group. Although the formal assessment of the student’s foreign language anxiety (FLA) was not conducted, the positive change in her behavior, which was observed by her teacher, supports the positive effect of high-immersion VR on FLA, confirming benefits of VR for students with initial high foreign language anxiety found by Thrasher (2022). Nevertheless, the fact that she chose the desktop because she did not want to be the only one using VR illustrates the reality of high-school social dynamics and peer pressure.

Finally, participants encountered a specific technical issue connected to technology accessibility, which was not emphasized in previous studies that reported technical problems

with using VR (Thrasher et al., 2024). The VR headsets would automatically go into “sleep mode” if participants had thick hair with bangs covering their forehead, which was a common issue among Latino participants.

### **Willingness to Communicate**

The current study explored how willingness to communicate changed throughout interactions with conversational AI based on participant ESL level and level of immersion. According to the WTC pre-test, initial students’ WTC is lower for ESL 1 level and slightly increases with each ESL level, with ESL 4 level showing the highest initial willingness to communicate. Descriptive statistics indicate that the overall effect of AI role-plays on growth in WTC was a 12 % increase for all ESL levels after three sessions and 19 % after five sessions total. These findings align with Tai & Chen (2020) and Ayedoun et al. (2019), who claim a positive effect of conversational AI on WTC. Due to the novelty of this technology, little research has explored how conversational AI in high-immersion VR benefits WTC. Previous findings showing increased engagement among students using VR (Xie et al., 2019), combined with positive perceptions of AI agents as fun (Moussalli & Cardoso, 2020) and motivational (Ebadi & Amini, 2022), should have led to higher WTC when these two technologies are combined. However, the current study showed that students using high-immersion VR had higher ESL levels, initially higher WTC, and showed less growth overall compared to those who preferred the desktop (RQ1).

The disagreement with previous findings could have occurred due to the preference for VR by 80 % of ESL 4 students, compared to just 22 % of ESL 1 level students. ESL 4 students participating in this study had a native-like level of oral proficiency, communicated with their Spanish-speaking peers in English, and were taking ESL classes to develop their written

communication skills. As their WTC in English was initially the highest among all levels, with a mean of 23.3 out of 30 points, it was expected that participants' WTC did not grow as fast as ESL 1 students', who had a mean of 15 out of 30 points initially.

### **Language production**

The impact of conversational AI on language production was explored as an average percentage growth in RQ2 and as the total number of syllables in RQ2.1 of this study. The average growth in language production for all participants was 49 %, which supports previous findings on the positive effect of language production (Dooly et al., 2023; Hassani et al., 2016; Hsu et al., 2021; Moussali & Cardoso, 2020). The Spearman correlation test indicated the highest negative correlation between the percentage growth in language production and ESL level (RQ2), but no significant correlation was found for language production as the total number of syllables during one full AI conversation and ESL level (RQ2.1). The findings indicate that lower-level participants produced more language over time with the average growth of 113 % for ESL 1 and 82 % for ESL 2 levels, while higher-level participants tended to have a negative change, - 7 % for ESL 3, or small growth of 7 % for ESL 4.

Language production growth also correlates with the usage of the desktop version of the platform instead of VR headsets. During classroom observations, high-level ESL students who used VR two times primarily engaged in conversations with each other rather than with AI bots. The affordances of VR, such as embodied cognition and presence, facilitated peer-to-peer interaction in VR (Karimi et al., 2023) by decreasing foreign language anxiety (Kaplan-Rakowski & Gruber, 2023). However, immersive conversational AI is currently represented by rather static interactions in the form of dialogue that does not encourage physical movements in a virtual environment. In other words, the affordance of VR for embodied cognition is not

integrated into conversations with AI agents. For instance, the environment where the conversation with AI takes place has a variety of interactive objects, but a conversation with an AI agent does not utilize the interactivity of the space. This study shows that once students switched from desktop to VR, their language production during AI conversations declined. Instead of conversing with AI bots, students were interacting with objects in the environment and their peers by accessing immersive environments together, even though they were instructed to enter it individually.

Another possible explanation for lower language production in a high-immersion environment as well as the evident preference of VR by high-level ESL students is connected to cognitive load and lower digital literacy for lower-level immigrant students. As research on cognitive load in a high-immersion environment was primarily conducted on university-level participants, only 10 % of studies in the review of Han et al. (2021), investigate how this technology affects the cognitive load of middle- and high-school students. Moreover, due to the higher cost of the technology, it is rarely integrated into underserved schools, where immigrant students from low-income households are usually clustered (Palardy et al., 2015). Some immigrant high-school students had their first *Chromebook* or tablet only at school in the U.S (Carhill-Poza & Williams, 2020), causing a “digital divide” in classrooms. It was also confirmed by one of the ESL teachers in this study, emphasizing the time it took to teach ESL students to use *Chromebooks* and the *Google Classroom* system used at the school. While a systematic review of research on cognitive load in VR does not provide a definite answer, it was found that students experience difficulty reading in virtual reality (Han et al., 2021), especially when the text is small. Even though there is limited research on the cognitive load of VR on immigrant

students with lower digital literacy, it can be stated that the lack of familiarity with technology could have discouraged students from using VR.

Interestingly, the spoken communication modality correlated with higher language production, but the average percentage growth in language production was not affected by modality. This was partially explained in RQ3, as students, who engaged in spoken conversations, experienced communication breakdowns caused by AI limited recognition of mispronounced words and low reliance on contextual cues. This is why lower-level students had a high percentage of repeated words and phrases in their overall language production. As some participants in this study were complete beginners in ESL, their repetition of words was also counted towards their total language produced. It explains a higher overall score for language production among students who preferred the spoken modality. However, the percentage growth of 59 % for the spoken modality is also substantially higher, than 14 % for the written modality. Even though participants experienced growth in language production over time, participants engaged in spoken conversations experienced a 45% higher growth than participants who preferred typing their responses. The findings of Payne and Whitney (2019), who claim that oral language production can be developed even by conversing in a written modality, confirm the overall positive effect of multimodality of conversational AI. However, the current study did not confirm their findings for a more positive effect of equal distribution of written and spoken communication for oral language production. Referring to the Complex Dynamic Systems Theory (CDST), this study confirms the dynamic nature of interaction with conversational AI. While students had a certain preference for one type of modality, 15 out of 36 students engaged in another type of communication, illustrating the importance of multimodal input possibilities, as 22% of students moved from writing to speaking over time.

## Individual differences in human-AI conversations

The current study investigates how immigrant students of different English proficiency levels communicate with AI. While lower-level ESL students tend to make use of embedded conversational features more than advanced ESL students (RQ2.2), the way students interacted with conversational AI not only differed for each individual but also evolved throughout the study with observed shifts in communication modality. More qualitative research needs to be done to explore less quantifiable patterns of communication, such as the authenticity of conversation, preference for a specific topic, and moments of communication breakdowns. Previous research on strategies during communication breakdown by Davis (2022) was extended by a possible explanation of moments when this communication breakdown occurs (RQ3). Limited recognition of mispronounced words was identified as one of the main patterns for communication breakdown, and it was also mentioned in numerous studies on different types of conversational AI (Chen et al., 2020; Dizon & Tang, 2020; Hsu et al., 2021; Moussalli & Cardoso, 2020). However, limited recognition of contextual cues was investigated in fewer studies. For instance, in the study of Tai (2022), which explored the effect of *Google Assistant* on 89 EFL Chinese students' speaking skills, one of the technical issues found was irrelevant responses of the IVA, which was also found in the current study in the form of limited recognition of contextual cues.

According to Katsarou et al.'s (2023) systematic review of voice-based Intelligent Virtual Agents (IVAs), this technology is particularly beneficial for intermediate-level students' oral skills and WTC development. IVAs are defined as one of eight types of conversational AI chatbots by Jeon and Choe (2023), emphasizing relevance of these studies for this novel technology. Findings that claim that advanced students benefit less from this technology and lack

authenticity of communication (Katsarou et al., 2023) were confirmed by this study due to stagnant progress in WTC and language production, as well as a negative change in language production, as well as the negative Spearman correlation with language production. However, evidence for lower-level ESL students not benefiting enough from this technology due to issues with comprehensibility (Katsarou et al., 2023) was not confirmed by this study. Even though students experienced communication breakdowns due to limited comprehension of mispronounced words, lower-level students who actively used the listening to word pronunciation feature had higher language production than other students. Compared to IVAs, such as *Siri* or *Google Assistant*, which were developed for everyday assistance in L1, the conversational AI developed by *Immerse* was programmed specifically for language learning with a set of features to support lower-level students' oral production.

## CHAPTER 6: CONCLUSION

The current study investigates the effect of immersive conversational AI on immigrant high-school students' willingness to communicate based on their ESL level and preferred level of immersion. It also explores how features and patterns of conversational AI influence language production depending on students' ESL level and preferred communication modality.

According to the findings in the study, lower-level ESL students prefer conversing with AI on desktop and have lower initial WTC. However, their growth in WTC was significantly higher compared to advanced ESL students who had higher initial WTC and were more likely to choose VR to converse with AI. (RQ1, pp. 30-37). It was also found that ESL 1 and ESL 2 students experience higher percentage growth in language production than more advanced ESL students. This growth also correlates with using the desktop version of the platform, but no significant correlation was found between growth in language production, WTC, and preferred communication modality (RQ2, pp. 37-40). Moreover, the conversational feature to listen to the pronunciation of a word correlates the most with higher language production, while an option to send a suggested response by AI correlated with lower language production (RQ2.1, pp. 40-42). However, all the conversational features are mostly used by lower-level students (RQ2.2, pp. 42-42). Finally, there were found two reoccurring patterns in human-AI communication disruption. The first one is limited recognition of mispronounced words, which mostly affected lower-level ESL students. The second pattern found referred to the limited recognition of contextual cues, which occurred in conversations of all students, leading to misinterpretation of the intended message (RQ3, pp. 42-43). These findings indicate that conversational AI can positively affect lower-level students' language production and WTC in L2 and can be integrated in curriculum to support the development of conversational skills of immigrant students.

## Limitations

This study has several limitations primarily associated with the novelty of the technology and the naturalistic setting of an underserved high school population where this technology was integrated. Firstly, the nature of ESL Resources classes, without a curriculum, grading procedures, or clear learning expectations, resulted in lower enthusiasm on the students' side to complete conversations with AI. During classroom observations, students sometimes could not complete role-plays as they had to retake their math or physics tests during ESL Resources class time. Students who did not have any homework or test commitments would often play online games, listen to Spanish music, or interact with their classmates in their L1. Therefore, role-play could have been perceived as a burden by some students, and teachers were not able to use grades as motivation for students to complete the conversations, which is not usually the case in a classroom environment.

Another limitation of the study is the incomplete dataset. As some students were absent and some could not or did not want to record their conversations, out of 30 data points for the WTC pre-test and 34 responses for WTC post-test 1, there were only 11 responses for WTC post-test, resulting in lower validity of findings that compare levels of immersion and willingness to communicate.

Finally, the diversity of data collected in the study with a rather small sample size makes inferential statistics inappropriate due to multiple violations of assumptions. Even though non-parametric statistics on the correlation between observed variables were conducted, linear regression could provide a deeper understanding not only of which variables correlate with WTC and language production but also to what extent each variable affects these two variables. The conducted Spearman correlation test was not as straightforward in determining whether higher-

level students have lower growth in WTC or students who use VR have lower growth in WTC, as students preferring VR are primarily advanced ESL students. However, in this case, a qualitative strand of video data of student interactions with AI complemented findings from the correlation test.

### **Implications**

As participants experienced higher growth in their language production and willingness to communicate, immersive conversational AI has the potential to be incorporated into school curriculums to develop conversational skills in a target language. However, several challenges encountered during study preparation and data collection lead to the following implications for the use of immersive conversational AI in a classroom.

In particular, the component of the study incorporating virtual reality required hours of equipment preparation, training, and consolidation of resources from researchers, school IT, administration, teachers, and students. The first step of setting up separate *Meta Quest* accounts for each participant took over forty minutes per participant. Even though it is highly recommended to use a headset management software (i.e., *Arbor XR*) to set up multiple devices on one account, it restricts the recording function embedded in *Meta Quest 2* headsets. As video data was the main data source for this study, separate *Meta Quest* accounts had to be created for each VR headset.

Additionally, logistical issues with transportation, secure headset storage in reorganized *Chromebook* Carts, and connecting all the headsets to the school Wi-Fi required collaboration with both school administration and IT personnel. Moreover, students experienced minor technical issues using VR headsets, making both teacher and student training essential for troubleshooting.

There were also some challenges associated with using desktops during the intervention. For instance, students did not want to put on headphones with a microphone and experienced issues with speech recognition because of that. Furthermore, screen recordings on desktops through the embedded *Chromebook* function were automatically encrypted, which made data cleaning more time-consuming. However, being present during data collection and getting to know all the participants was essential in identifying recordings.

Overall, there is great potential for immersive conversational AI to empower students to practice second language production. Nevertheless, integrating technology into the classroom requires thorough planning and preparation, as well as establishing rapport with all the stakeholders.

### **Future Research**

Adding to the pool of research on immersive conversational AI, this study showcases how crucial it is to investigate the role of new technological advances on the immigrant high school population in underserved American schools. As the number of studies combining high-immersion virtual reality and conversational AI is highly limited, it is crucial to continue conducting research on the intersection of extended reality (XR) and conversational AI, especially in naturalistic settings with a higher number of participants. Moreover, the role of teachers in the integration of immersive conversational AI should be explored more deeply, as they play a crucial role in both facilitating this integration and limiting it.

As generative AI is constantly evolving, the found limitations of speech recognition might be less apparent in the newer version GPT-4o. Therefore, there is a need for consistent research on utilizing generative AI for language learning and speaking skills development to make sure that research is representative of the conversational AI capabilities.

## REFERENCES

- Alefesha, H. M. N., & Al-Jamal, D. A. H. (2019). Syrian Refugees' Challenges and Problems of Learning and Teaching English as a Foreign Language (EFL): Jordan as an Example. *Journal of Ethnic and Cultural Studies*, 6(1), 117–129. <https://doi.org/10.29333/ejecs/235>
- Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding Communicative and Affective Strategies to an Embodied Conversational Agent to Enhance Second Language Learners' Willingness to Communicate. *International Journal of Artificial Intelligence in Education*, 29(1), 29–57. <https://doi.org/10.1007/s40593-018-0171-6>
- Bolter, J. D. (1984). Artificial Intelligence. *Daedalus*, 113(3), 1–18. <http://www.jstor.org/stable/20024925>
- Carhill-Poza, A., & Williams, T. P. (2020). Learning “Anytime, Anywhere”? The Imperfect Alignment of Immigrant Students' Experiences and School-Based Technologies in an Urban US High School. *Comparative Education Review*, 64(3), 428–450. <https://doi.org/10.1086/709430>
- Chen, H. H. J., Yang, C. T. Y., & Lai, K. K. W. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350. <https://doi.org/10.1080/10494820.2020.1833043>
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, 25(1), 28–47. [https://doi.org/10.30191/ETS.202201\\_25\(1\).0003](https://doi.org/10.30191/ETS.202201_25(1).0003)

- Clasen, D. R., & Brown, B. B. (1985). The multidimensionality of peer pressure in adolescence. *Journal of Youth and Adolescence*, 14(6), 451–468. <https://doi.org/10.1007/BF02139520>
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd Ed.). Sage Publications.
- Davis, G. M. (2022). *Humanlike Conversational Artificial Intelligence Agents for Foreign or Second Language Learning: User Perceptions, Expectations, And Interactions with Agents* [Ph.D., Stanford University].  
<https://www.proquest.com/pqdtglobal/docview/2723858200/abstract/474BA81BEE284C85PQ/1>
- DeCapua, A., & Marshall, H. W. (2011). Reaching ELLs at Risk: Instruction for Students With Limited or Interrupted Formal Education. *Preventing School Failure: Alternative Education for Children and Youth*, 55(1), 35–41.  
<https://doi.org/10.1080/10459880903291680>
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-base perspectives for L2 teaching and research*. John Benjamins North America.
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332–2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *The JALT CALL Journal*, 16(2), 107–120.  
<https://doi.org/10.29140/jaltcall.v16n2.273>

- Dooly, M., Thrasher, T., & Sadler, R. (2023). “Whoa! Incredible!” Language Learning Experiences in Virtual Reality. *RELC Journal*, 00336882231167610.  
<https://doi.org/10.1177/00336882231167610>
- Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., Janssen, M., ... Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 66, 102542.  
<https://doi.org/10.1016/j.ijinfomgt.2022.102542>
- Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners’ motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive learning environments*.  
<https://doi.org/10.1080/10494820.2022.2096638>
- Edwards, M., Adams, E. M., Waldo, M., Hadfield, O. D., & Biegel, G. M. (2014). Effects of a Mindfulness Group on Latino Adolescent Students: Examining Levels of Perceived Stress, Mindfulness, Self-Compassion, and Psychological Symptoms. *The Journal for Specialists in Group Work*, 39(2), 145–163.  
<https://doi.org/10.1080/01933922.2014.891683>
- Gallagher, H. C. (2013). Willingness to Communicate and Cross-cultural Adaptation: L2 Communication and Acculturative Stress as Transaction. *Applied Linguistics*, 34(1), 53–73. <https://doi.org/10.1093/applin/ams023>

- Gass, S. M., & Varonis, E. M. (1994). Input, Interaction, and Second Language Production. *Studies in Second Language Acquisition*, 16(3), 283–302.  
<https://doi.org/10.1017/S0272263100013097>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Hsu, H.-L., Chen, H. H.-J., & Todd, A. G. (2021). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, 1–14. <https://doi.org/10.1080/10494820.2021.2016864>
- Huang, X., Zou, D., Cheng, G., & Xie, H. (2021). A systematic review of AR and VR enhanced language learning. *Sustainability*, 13(4639), 1-28. <https://doi.org/10.3390/su13094639>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301\\_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. <https://doi.org/10.1016/j.compedu.2023.104898>
- Kanno, Y., & Varghese, M. M. (2010). Immigrant and Refugee ESL Students' Challenges to Accessing Four-Year College Education: From Language Policy to Educational Policy. *Journal of Language, Identity & Education*, 9(5), 310–328.  
<https://doi.org/10.1080/15348458.2010.517693>
- Karimi, H., Sañosa, D. J., Hernandez Rios, K., Tran, P., Chun, D. M., Wang, R., & Arya, D. J. (2023). Building a City in the Sky: Multiliteracies in Immersive Virtual Reality. *CALICO Journal*, 40(1), 24–44. <https://doi.org/10.1558/cj.22838>

- 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.
- Kaplan-Rakowski, R., & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety. *Smart Learning Environments*, 10(1), 46.  
<https://doi.org/10.1186/s40561-023-00263-9>
- Katsarou, E., Wild, F., Sougari, A.-M., & Chatzipanagiotou, P. (2023). A Systematic Review of Voice-based Intelligent Virtual Agents in EFL Education. *International Journal of Emerging Technologies in Learning (Online)*, 18(10), 65–85.  
<https://doi.org/10.3991/ijet.v18i10.37723>
- Kramersch, C., & Steffensen, S. (2008). Ecological perspectives on second language acquisition and socialization. In P. A. Duff, & N. H. Hornberger (Eds), *Encyclopedia of Language and Education* (2nd ed), 17–28. New York: Springer. [https://doi.org/10.1007/978-0-387-30424-3\\_194](https://doi.org/10.1007/978-0-387-30424-3_194)
- Krashen, S. D. (1982). *Principles and practice in second language acquisition* (1st ed). Pergamon.
- Larsen-Freeman, D. (2018). Looking ahead: Future directions in, and future research into, second language acquisition. *Foreign Language Annals*, 51(1), 55–72.  
<https://doi.org/10.1111/flan.12314>
- Lee, J. S., & Lee, K. (2020). Role of L2 Motivational Self System on Willingness to Communicate of Korean EFL University and Secondary Students. *Journal of Psycholinguist Research*, 49, 147–161. <https://doi.org/10.1007/s10936-019-09675-6>

- Li, X., & Grineva, M. (2017). Academic and Social Adjustment of High School Refugee Youth in Newfoundland. *TESL Canada Journal*, 34(1), 51–71.  
<https://doi.org/10.18806/tesl.v34i1.1255>
- Lu, S., Huang, C., Cheung, S. P., Rios, J. A., & Chen, Y. (2021). Mindfulness and social-emotional skills in Latino pre-adolescents in the U.S.: The mediating role of executive function. *Health & Social Care in the Community*, 29(4), 1010–1018.  
<https://doi.org/10.1111/hsc.13135>
- Macintyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing Willingness to Communicate in a L2: A Situational Model of L2 Confidence and Affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Mirzaei, M. S., Zhang, Q., Van Der Struijk, S., & Nishida, T. (2018). Language learning through conversation envisioning in virtual reality: A sociocultural approach. In P. Taalas, J. Jalkanen, L. Bradley, & S. Thouésny, *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 207–213).  
<https://doi.org/10.14705/rpnet.2018.26.838>
- Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, 33(8), 865–890. <https://doi.org/10.1080/09588221.2019.1595664>
- Mroz, A. (2015). The development of second language critical thinking in a virtual language learning environment: A process-oriented mixed-method study. *CALICO Journal*, 32(3), 528–553. <https://doi.org/10.1558/cj.v32i3.26386>

- Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38(3), 289-306.  
<https://doi.org/10.1177/002383099503800305>
- Ortiz, J., Smith, B., Shelley, B., & Erickson, K. (2019). Adapting Mindfulness to Engage Latinos and Improve Mental Health in Primary Care: A Pilot Study. *Mindfulness*, 10.  
<https://doi.org/10.1007/s12671-019-01229-0>
- Palardy, G., Rumberger, R., & Butler, T. (2015). The Effect of High School Socioeconomic, Racial, and Linguistic Segregation on Academic Performance and School Behaviors. *Teachers College Record: The Voice of Scholarship in Education*, 117(12), 1–52.  
<https://doi.org/10.1177/016146811511701206>
- Payne, J. S., & Whitney, P. J. (2019). Developing L2 Oral Proficiency through Synchronous CMC: Output, Working Memory, and Interlanguage Development. *CALICO Journal*, 20(1), 7–32. <https://doi.org/10.1558/cj.v20i1.7-32>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(4), 329-347.  
<https://doi.org/10.1017/S0140525X12001495>
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-Powered Conversational System for Second Language Learning. *26th International Conference on Intelligent User Interfaces*, 434–444.  
<https://doi.org/10.1145/3397481.3450648>
- Sadler, R. W. (2020). Virtual Landscapes. In M. Dressman, & R. W. Sadler (Eds.), *The Handbook of Informal Language Learning* (pp. 85-100). Wiley-Blackwell. <https://doi.org/10.1002/9781119472384.ch6>

- Tai, A. D. (2017). *Exploring the Dynamics of Willingness to Communicate in Written Communication: A Case Study*. 17(1), 39–55. <https://doi.org/10.7916/D8612B78>
- Tai, T.-Y., & Chen, H. H.-J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*. Scopus. <https://doi.org/10.1080/10494820.2020.1841801>
- Tai, T. Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2022.2075013>
- Thrasher, T. (2022). The Impact of Virtual Reality on L2 French Learners' Language Anxiety and Oral Comprehensibility: An Exploratory Study. *CALICO Journal*, 39(2), 219–238. <https://doi.org/10.1558/cj.42198>
- Thrasher, T., Kaplan-Rakowski, R., Ovsianikova, U., Meyr, J., & Yuan, Y. (2024). "I Can Talk to Spanish Speakers in Illinois!": Student Perspectives on AI-Avatar Role Plays in Virtual Reality. 157–165. <https://doi.org/10.22492/issn.2759-1182.2024.18>
- Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: a study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, 37(4), 814–840. <https://doi.org/10.1080/09588221.2022.2056203>
- Woo, J. H., & Choi, H. (2021). Systematic Review for AI-based Language Learning Tools. *Journal of Digital Contents Society*, 22(11), 1783–1792. <https://doi.org/10.9728/dcs.2021.22.11.1783>

Xie, Y., Ryder, L. & Chen, Y. Using Interactive Virtual Reality Tools in an Advanced Chinese Language Class: a Case Study. *TechTrends* 63, 251–259 (2019).

<https://doi.org/10.1007/s11528-019-00389-z>

Yang, H., Kim, H., Lee, J., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343.

<https://doi.org/10.1017/S0958344022000039>

## APPENDIX A: INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL



Office of the Vice Chancellor for Research & Innovation

Office for the Protection of Research Subjects  
1901 S. First St., Suite A, MC-685  
Champaign, IL 61820

### Notice of Approval: Amendment 02

September 4, 2023

<b>Principal Investigator</b>	Randall Sadler
<b>CC</b>	Uliana Ovsianikova, Tricia Thrasher
<b>Protocol Title</b>	Virtual Reality for Education (VR4Ed)
<b>Protocol Number</b>	23789
<b>Funding Source</b>	Joint gift from Meta and Immerse
<b>Review Type</b>	Expedited 6, 7
<b>Amendment Requested</b>	Adding Uliana Ovsianikova to the research team
<b>Status</b>	Active
<b>Risk Determination</b>	No greater than minimal risk
<b>Amendment Approval Date</b>	September 4, 2023
<b>Expiration Date</b>	April 12, 2024

This letter authorizes the use of human subjects in the above protocol. The University of Illinois at Urbana-Champaign Institutional Review Board (IRB) has reviewed and approved the research study as described.

The Principal Investigator of this study is responsible for:

- Conducting research in a manner consistent with the requirements of the University and federal regulations found at 45 CFR 46.
- Using the approved consent documents, with the footer, from this approved package.
- Requesting approval from the IRB prior to implementing modifications.
- Notifying OPRS of any problems involving human subjects, including unanticipated events, participant complaints, or protocol deviations.
- Notifying OPRS of the completion of the study.

## APPENDIX B: LANGUAGE AND SOCIAL BACKGROUND QUESTIONNAIRE

Q1 ¿Cuál es tu nombre?

---

Q2 ¿Cuántos años tienes?

---

Q3 ¿Cuál es tu género?

- Masculino (1)
- Femenino (2)
- Otro (3)

Q4 ¿Cuál es tu nacionalidad?

---

Q5 ¿Cuánto tiempo has vivido en los Estados Unidos?

- Menos de 6 meses (1)
- 6 meses-1 año (2)
- 1 año-2 años (3)
- Más de 2 años (4)

Q6 ¿Cuánto tiempo has estudiado inglés?

- Menos de 6 meses (1)
- 6 meses-1 año (2)
- 1 año-2 años (3)
- Más de 2 años (4)

Q7 ¿En qué grado estás en la escuela?

---

Q8 ¿Cuál es tu nivel de inglés? (Si no sabes, pregunta a tu profesora).

- ESL 1 (1)
- ESL 2 (2)
- ESL 3 (3)
- ESL 4 (4)
- Other (5) \_\_\_\_\_

Q9 ¿Qué idiomas conoces? (por “conocer” me refiero a la capacidad de tener al menos un intercambio básico en el idioma).

- Español (1)
- Inglés (2)
- Otro (3) \_\_\_\_\_

**End of Block: Cuestionario de idioma y contexto social**

---

**Start of Block: Cuestionario de idioma y contexto social: familia**

Q10 ¿Cuál es la ocupación de tu madre?

---

Q11 ¿Qué idiomas conoce tu madre?

- Español (1)
- Inglés (2)
- Otro (3) \_\_\_\_\_

---

Q12 ¿Cuál es la ocupación de tu padre?

---

Q13 ¿Qué idiomas conoce tu padre?

- Español (1)
- Inglés (2)
- Otro (3) \_\_\_\_\_

---

Q14 Indique el nivel de educación más alto y ocupación de cada uno de los padres:

	Sin diploma de escuela secundaria (1)	Diploma de escuela secundaria (2)	Algo de educación postsecundaria (3)	Título o diploma postsecundario (4)	Título de posgrado o profesional (5)
Madre (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Padre (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

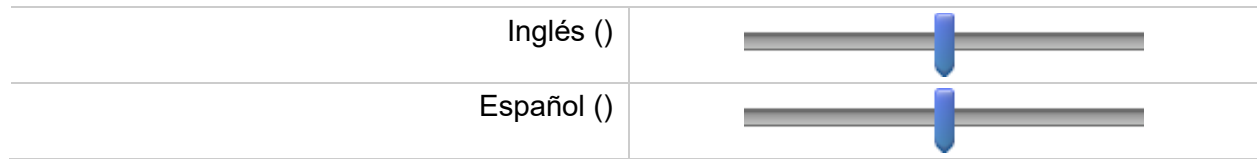
---

Q15 Por favor, califica las siguientes afirmaciones en una escala desde "Muy desfavorable" hasta "Muy favorable."

	Muy desfavorable (1)	Desfavorable (2)	Neutral (3)	Favorable (4)	Muy favorable (5)
¿Cuál es tu actitud hacia el idioma inglés? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
¿Cuál es tu actitud hacia tu profesor de inglés? (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q16 Con qué frecuencia tu profesora utiliza los idiomas en ESL Resource clase?

0 10 20 30 40 50 60 70 80 90 100



Q17 ¿Cómo describirías tu desempeño en inglés en comparación con el resto del grupo?

- Muy por debajo del promedio (1)
- Por debajo del promedio (2)
- Promedio (3)
- Por encima del promedio (4)
- Muy por encima del promedio (5)

Q18 ¿Te describirías en inglés como:

- Principiante (1)
- Intermedio bajo (2)
- Intermedio (3)
- Intermedio alto (4)
- Avanzado (5)

Q19 Piensa en cómo actuarías en inglés. ¿Qué posibilidades hay de que INICIES una conversación? Marca la respuesta más adecuada.

	Casi nunca dispuesto (1)	A veces dispuesto (2)	Dispuesto la mitad del tiempo (3)	Generalmente dispuesto (4)	Casi siempre dispuesto (5)
Hablar en grupo sobre tus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

vacaciones  
de verano (1)

Hablar con tu  
profesor  
sobre tu tarea  
(2)

Un extraño  
entra a la  
habitación en  
la que estás,  
¿Qué tan  
dispuesto  
estarías a  
tener una  
conversación  
si él hablara  
contigo  
primero? (3)

Estás  
confundido  
acerca de  
una tarea que  
debes  
completar,  
¿Qué tan  
dispuesto  
estás a pedir  
instrucciones/  
aclaraciones?  
(4)

Hablar con  
un amigo  
durante la  
clase. (5)

¿Qué tan  
dispuesto  
estarías a ser  
actor en una  
obra de  
teatro? (6)

Describe las  
reglas de tu  
juego  
favorito. (7)



Juega un  
juego en  
inglés (8)



End of Block: Cuestionario de idioma y contexto social: familia

Start of Block: Antecedentes del idioma

Q20 En relación con el desempeño de un hablante altamente competente, califica tu nivel de competencia en una escala de 0 a 100% para las siguientes actividades realizadas en **INGLÉS**

0 10 20 30 40 50 60 70 80 90 100







Q21 Del tiempo que dedicas a cada una de las siguientes actividades, ¿cuánto de ese tiempo lo realizas en **INGLÉS** ?

	Ningún (1)	Poco (2)	Algún (3)	Mayoría (4)	Todo (5)
Hablando (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escuchando (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leyendo (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escribiendo (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q22 En relación con el desempeño de un hablante altamente competente, califica tu nivel de competencia en una escala de 0 a 100% para las siguientes actividades realizadas en **ESPAÑOL**

0 10 20 30 40 50 60 70 80 90 100

Expresión oral ( )	
Comprensión ( )	
Comprensión lectora ( )	
Expresión escrita ( )	

Q23 Del tiempo que dedicas a cada una de las siguientes actividades, ¿cuánto de ese tiempo lo realizas en **ESPAÑOL**?

	Ningún (1)	Poco (2)	Algún (3)	Mayoría (4)	Todo (5)
Hablando (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escuchando (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leyendo (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escribiendo (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Antecedentes del idioma

Start of Block: Comportamiento de uso del lenguaje comunitario

Q24 Indica qué lengua(s) escuchaste o utilizaste con más frecuencia en las siguientes etapas de la vida, tanto dentro como fuera de casa.

	Todo inglés (1)	Mayormente inglés (2)	Mitad inglés mitad español (3)	Mayormente español (4)	Solo español (5)
Infancia (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Edad preescolar (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Edad de escuela primaria (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Edad de secundaria (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q25 Indica qué idioma(s) utilizas generalmente cuando hablas con las siguientes personas.

	Todo inglés (1)	Mayormente inglés (2)	Mitad inglés mitad español (3)	Mayormente español (4)	Solo español (5)
Padres (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hermanos (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Otros parientes (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Amigos (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compañeros de clase (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

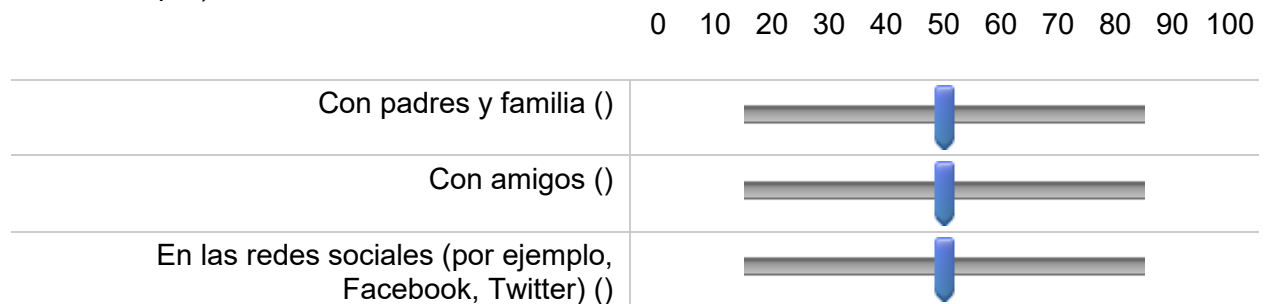
Q26 Indica qué idioma(s) utilizas generalmente en las siguientes situaciones

	Todo inglés (1)	Mayormente inglés (2)	Mitad inglés mitad español (3)	Mayormente español (4)	Solo español (5)
Hogar (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Escuela (11)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trabajo (12)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Actividades sociales (por ejemplo, salir con amigos, ver películas) (13)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Actividades extracurriculares (por ejemplo, pasatiempos, deportes, voluntariado, juegos) (14)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compras/ Restaurantes/ Otros servicios comerciales (15)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q27 Indica qué idioma(s) utilizas generalmente para las siguientes actividades.

	Todo inglés (1)	Mayormente inglés (2)	Mitad inglés mitad español (3)	Mayormente español (4)	Solo español (5)
Lectura (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Envío de mensajes de texto (16)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Redes sociales (por ejemplo, Facebook, Twitter, etc.) (17)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Redacción de listas de compras, notas, etc. (18)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ver películas (19)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Navegación por Internet (20)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q28 Algunas personas cambian entre los idiomas que conocen en una sola conversación (es decir, mientras hablan en un idioma pueden usar oraciones o palabras del otro idioma). Esto se conoce como "cambio de idioma". Indica con qué frecuencia cambias de idioma (0-nunca; 100%-siempre)



## APPENDIX C: WILLINGNESS TO COMMUNICATE QUESTIONNAIRE



Q1 ¿Cuál es tu nombre? \_\_\_\_\_

Q2 Por favor, califica las siguientes afirmaciones en una escala desde "Muy desfavorable" hasta "Muy favorable."

	Muy desfavorable (1)	Desfavorable (2)	Neutral (3)	Favorable (4)	Muy favorable (5)
¿Cuál es tu actitud hacia el idioma inglés ? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
¿Cuál es tu actitud hacia tu profesor de inglés? (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q3 ¿Con qué frecuencia tu profesora utiliza los idiomas en ESL Resource clase?

0 10 20 30 40 50 60 70 80 90 100

Inglés ( )	
Español ( )	

Q4 ¿Cómo describirías tu desempeño en inglés en comparación con el resto del grupo?

- Muy por debajo del promedio (1)
- Por debajo del promedio (2)
- Promedio (3)
- Por encima del promedio (4)
- Muy por encima del promedio (5)

Q5 ¿Te describirías en inglés como:

- Principiante (1)
- Intermedio bajo (2)
- Intermedio (3)
- Intermedio alto (4)
- Avanzado (5)

Q6 Piensa en cómo actuarías en inglés. ¿Qué posibilidades hay de que INICIES una conversación? Marca la respuesta más adecuada.

	Casi nunca dispuesto (1)	A veces dispuesto (2)	Dispuesto la mitad del tiempo (3)	Generalmente dispuesto (4)	Casi siempre dispuesto (5)
Hablar en grupo sobre tus vacaciones de verano (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hablar con tu profesor sobre tu tarea (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Un extraño entra a la habitación en la que estás, ¿Qué tan dispuesto estarías a tener una conversación si él hablara contigo primero? (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Estás confundido acerca de una tarea que debes completar, ¿Qué tan dispuesto estás a pedir instrucciones/ aclaraciones? (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Hablar con un amigo durante la clase. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
¿Qué tan dispuesto estarías a ser actor en una obra de teatro? (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Describe las reglas de tu juego favorito. (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Juega un juego en inglés (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## APPENDIX D: PYTHON CODES FOR SPEARMAN CORRELATIONS

```
"""Calculates individual correlations between WTC, ESL level, and level of
immersion"""

import pandas as pd
from scipy.stats import spearmanr

data = pd.read_csv('data/wtc_level.csv')
def esl_wtc_pre():
    data.dropna(subset=['ESL level', 'Pre-test'], inplace=True)
    correlation, p_value = spearmanr(data['ESL level'], data['Pre-test'])
    print(f'\nCorrelation between ESL level and WTC pre-test: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
esl_wtc_pre()

def esl_wtc_post_1():
    data.dropna(subset=['ESL level', 'Post-test 1'], inplace=True)
    correlation, p_value = spearmanr(data['ESL level'], data['Post-test 1'])
    print(f'\nCorrelation between ESL level and WTC post-test 1: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
esl_wtc_post_1()

def esl_wtc_post_2():
    data.dropna(subset=['ESL level', 'Post-test 2'], inplace=True)
    correlation, p_value = spearmanr(data['ESL level'], data['Post-test 2'])
    print(f'\nCorrelation between ESL level and WTC post-test 2: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
esl_wtc_post_2()

def esl_wtc_growth_post_1():
    data.dropna(subset=['ESL level', 'Percentage Growth (Pre-test vs Post-
test 1)'], inplace=True)
    correlation, p_value = spearmanr(data['ESL level'], data['Percentage
Growth (Pre-test vs Post-test 1)'])
    print(f'\nCorrelation between ESL level and Percentage Growth (Pre-test
vs Post-test 1): ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
esl_wtc_growth_post_1()

def esl_wtc_growth_post_2():
    data.dropna(subset=['ESL level', 'Percentage Growth (Pre-test vs Post-
test 2)'], inplace=True)
    correlation, p_value = spearmanr(data['ESL level'], data['Percentage
Growth (Pre-test vs Post-test 2)'])
    print(f'\nCorrelation between ESL level and Percentage Growth (Pre-test
vs Post-test 2): ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
esl_wtc_growth_post_2()
```

```

def level_wtc_pre():
    data.dropna(subset=['level_immersion', 'Pre-test'], inplace=True)
    correlation, p_value = spearmanr(data['level_immersion'], data['Pre-test'])
    print(f'\nCorrelation between level of immersion and WTC pre-test: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
level_wtc_pre()

def level_wtc_post_1():
    data.dropna(subset=['level_immersion', 'Post-test 1'], inplace=True)
    correlation, p_value = spearmanr(data['level_immersion'], data['Post-test 1'])
    print(f'\nCorrelation between level of immersion and WTC post-test 1: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
level_wtc_post_1()

def level_wtc_pre_post_1_growth():
    data.dropna(subset=['level_immersion', 'Percentage Growth (Pre-test vs Post-test 1)'], inplace=True)
    correlation, p_value = spearmanr(data['level_immersion'], data['Percentage Growth (Pre-test vs Post-test 1)'])
    print(f'\nCorrelation between level of immersion and % change in WTC from pre-test to post-test 1: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
level_wtc_pre_post_1_growth()

def level_wtc_post_2():
    data.dropna(subset=['level_immersion', 'Post-test 2'], inplace=True)
    correlation, p_value = spearmanr(data['level_immersion'], data['Post-test 2'])
    print(f'\nCorrelation between level of immersion and WTC post-test 2: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
level_wtc_post_2()

def level_wtc_pre_post_2_growth():
    data.dropna(subset=['level_immersion', 'Percentage Growth (Pre-test vs Post-test 2)'], inplace=True)
    correlation, p_value = spearmanr(data['level_immersion'], data['Percentage Growth (Pre-test vs Post-test 2)'])
    print(f'\nCorrelation between level of immersion and % change in WTC from pre-test to post-test 2: ')
    print("Spearman correlation coefficient:", correlation)
    print("p-value:", p_value)
level_wtc_pre_post_2_growth()

```

```

"""Visualizes Spearman Coefficients' and p-value for WTC, ESL level, and level of immersion"""
import matplotlib.pyplot as plt
variable_names = ["ESL level vs. WTC pre-test", "ESL level vs. WTC post-test 1",
                  "ESL level vs. WTC post-test 2",
                  "ESL level vs. % Growth (Pre vs. Post 1)", "ESL level vs. % Growth (Pre vs. Post 2)",

```

```

        "Level of immersion vs. WTC pre-test",
        "Level of immersion vs. WTC post-test 1", "Level of
immersion vs. % Growth (Pre vs. Post 1)",
        "Level of immersion vs. WTC post-test 2",
        "Level of immersion vs. % Growth (Pre vs. Post 2)"]
p_values = [0.038, 0.294, 0.678, 0.547, 0.759, 0.629, 0.628, 0.545, 0.466,
0.631]
correlation_coefficients = [0.380, 0.206, 0.151, 0.217, 0.111, -0.175, -
0.175, -0.218, -0.261, -0.174]
bubble_sizes = [abs(coeff) * 7000 for coeff in correlation_coefficients]

plt.figure(figsize=(12, 8))
for i in range(len(variable_names)):
    if p_values[i] <= 0.05:
        p_color = '#ED7D31'
    elif 0.05 < p_values[i] <= 0.2:
        p_color = '#ED7D31'
    elif 0.2 < p_values[i] <= 0.5:
        p_color = '#D56E22'
    else:
        p_color = '#D56E22'

    name_parts = variable_names[i].split(" vs. ")
    name = "\nvs. ".join(name_parts)

    if "Level of immersion" in variable_names[i] and "%" in
variable_names[i]:
        plt.scatter(p_values[i], i, s=bubble_sizes[i], marker='o',
color=p_color, edgecolor='black', linewidth=1.5,
alpha=0.8)
        plt.text(p_values[i], i - 0.6, f'{correlation_coefficients[i]:.3f}',
ha='center', va='top', fontsize=12,
fontweight='bold', color='black')
    elif "Level of immersion" in variable_names[i] and "WTC pre-test" in
variable_names[i]:
        plt.scatter(p_values[i], i, s=bubble_sizes[i], marker='o',
color=p_color, edgecolor='black', linewidth=1.5,
alpha=0.8)
        plt.text(p_values[i], i - 0.6, f'{correlation_coefficients[i]:.3f}',
ha='center', va='top', fontsize=12,
fontweight='bold', color='black')
    elif correlation_coefficients[i] < 0:
        plt.scatter(p_values[i], i, s=bubble_sizes[i], marker='o',
color=p_color, edgecolor='black', linewidth=1.5,
alpha=0.8)
        plt.text(p_values[i], i + 0.5, f'-
{abs(correlation_coefficients[i]):.3f}', ha='center', va='bottom',
fontweight='bold', color='black')
    else:
        plt.scatter(p_values[i], i, s=bubble_sizes[i], marker='o',
color='#4F81BD', edgecolor='black', linewidth=1.5,
alpha=0.8)
        plt.text(p_values[i], i + 0.5, f'{correlation_coefficients[i]:.3f}',
ha='center', va='bottom', fontsize=12,
fontweight='bold', color='black')

plt.xticks([0, 0.1, 0.2, 0.5, 0.7])

```

```

plt.yticks(range(len(variable_names)), [name for name in variable_names],
           fontsize=12)
plt.axvline(x=0.05, color='red', linestyle='--', linewidth=1.5)
plt.xlabel('p-values', fontsize=12)
plt.ylabel('Variable Names', fontsize=12)
plt.title('Bubble Chart of Correlation Coefficients and p-values',
         fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()

```

```

"""Calculates Spearman correlation matrix for average percentage growth in
language production"""
import pandas as pd
from scipy.stats import spearmanr

def spear_test():
    data = pd.read_csv('data/lang_prod_growth.csv')
    data.dropna(subset=['lang_prod_growth_av', 'esl_level',
'level_immersion', 'modality', 'wtc_pre', 'wtc_growth'], inplace=True)

    corr_matrix, p_values = spearmanr(data[['lang_prod_growth_av',
'esl_level', 'level_immersion', 'modality', 'wtc_pre', 'wtc_growth']])
    columns = ['lang_prod_growth_av', 'esl_level', 'level_immersion',
'modality', 'wtc_pre', 'wtc_growth']

    corr_df = pd.DataFrame(corr_matrix, index=columns, columns=columns)
    p_values_df = pd.DataFrame(p_values, index=columns, columns=columns)

    pd.set_option('display.max_columns', None)

    print("Spearman Correlation Matrix:")
    print(corr_df)

    print("\nP-Values Matrix:")
    print(p_values_df)

spear_test()

```

```

"""Visualizes Spearman Correlation Matrix for average percentage growth in
language production"""
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

corr_matrix_updated = pd.DataFrame({
    'Lang Prod % Growth': [1.000000, -0.421621, 0.397919, -0.045948, -
0.327931, 0.009882],
    'ESL Level': [-0.421621, 1.000000, -0.674269, 0.047668, 0.538546, -
0.257582],
    'Level of Immersion': [0.397919, -0.674269, 1.000000, 0.082479, -
0.461844, 0.062086],
    'Modality': [-0.045948, 0.047668, 0.082479, 1.000000, 0.164090, -
0.368693],

```

```

    'WTC Pre': [-0.327931, 0.538546, -0.461844, 0.164090, 1.000000, -
0.564930],
    'WTC Growth': [0.009882, -0.257582, 0.062086, -0.368693, -0.564930,
1.000000]
}, index=['Lang Prod % Growth', 'ESL Level', 'Level of Immersion',
'Modality', 'WTC Pre', 'WTC Growth'])

pval_matrix_updated = pd.DataFrame({
    'Lang Prod % Growth': [4.049407e-201, 0.025441, 0.035987, 0.816406,
0.088454, 0.960196],
    'ESL Level': [0.025441, 0.000000, 0.000083, 0.809654, 0.003111,
0.185731],
    'Level of Immersion': [0.035987, 0.000083, 0.000000, 0.676496, 0.013357,
0.753634],
    'Modality': [0.816406, 0.809654, 0.676496, 0.000000, 0.404072, 0.053529],
    'WTC Pre': [0.088454, 0.003111, 0.013357, 0.404072, 0.000000, 0.001735],
    'WTC Growth': [0.960196, 0.185731, 0.753634, 0.053529, 0.001735,
0.000000]
}, index=['Lang Prod % Growth', 'ESL Level', 'Level of Immersion',
'Modality', 'WTC Pre', 'WTC Growth'])

mask_updated = np.triu(np.ones_like(corr_matrix_updated, dtype=bool))
combined_corr_matrix = corr_matrix_updated.mask(mask_updated)

plt.figure(figsize=(10, 8))

sns.heatmap(combined_corr_matrix, annot=True, fmt=".2f",
cmap=sns.diverging_palette(220, 20, as_cmap=True), center=0, linewidths=.5,
vmin=-1, vmax=1, square=True, cbar_kws={'label': 'Spearman
Correlation (r)'})

for i in range(combined_corr_matrix.shape[0]):
    for j in range(combined_corr_matrix.shape[1]):
        if i != j and not np.isnan(combined_corr_matrix.iloc[i, j]):
            r_value = combined_corr_matrix.iloc[i, j]
            plt.text(j + 0.5, i + 0.5, f'{r_value:.2f}', ha='center',
va='center', color='black', fontsize=14)

plt.xticks(ticks=np.arange(len(combined_corr_matrix.columns)) + 0.5,
labels=combined_corr_matrix.columns, rotation=45, ha='right', fontsize=14)
plt.yticks(ticks=np.arange(len(combined_corr_matrix.index)) + 0.5,
labels=combined_corr_matrix.index, rotation=0, fontsize=14)
plt.title('Spearman Correlation Matrix', fontsize=16)
plt.tight_layout()
plt.show()

```

```

"""Calculates Spearman Correlation Matrix with Spearman coefficients and p-
values"""
import pandas as pd
from scipy.stats import spearmanr
def spear_test():
    data = pd.read_csv('data/sessions.csv')

    data.dropna(subset=['lang_prod', 'esl_level', 'level_of_immersion',
'modality', 'hints', 'trans_phr', 'def_word',

```

```

        'list_phr', 'list_word', 'send_hint', 'cor_wr',
'google'], inplace=True)

    corr_matrix, p_values = spearmanr(data[['lang_prod', 'esl_level',
'level_of_immersion', 'modality', 'hints', 'trans_phr', 'def_word',
        'list_phr', 'list_word', 'send_hint', 'cor_wr',
'google']])

    columns = ['lang_prod', 'esl_level', 'level_of_immersion',
'modality', 'hints', 'trans_phr', 'def_word',
        'list_phr', 'list_word', 'send_hint', 'cor_wr',
'google']

    corr_df = pd.DataFrame(corr_matrix, index=columns, columns=columns)
    p_values_df = pd.DataFrame(p_values, index=columns, columns=columns)

    pd.set_option('display.max_columns', None)

    print("Spearman Correlation Matrix:")
    print(corr_df)

    print("\nP-Values Matrix:")
    print(p_values_df)
spear_test()

```

```

"""This code visualizes Spearman Correlation Matrix showing Spearman
coefficients"""

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

def visualize_spearman_matrix(corr_df, p_values_df, labels):
    significance_mask = ((corr_df > 0.1) | (corr_df < -0.1))
    filtered_corr = corr_df.where(significance_mask, other=0)
    cmap = sns.diverging_palette(220, 20, as_cmap=True)
    mask = np.triu(np.ones_like(filtered_corr, dtype=bool))
    plt.figure(figsize=(12, 10))
    sns.heatmap(filtered_corr, annot=True, mask=mask, cmap=cmap, fmt=".2f",
linewidths=0.5,
                cbar_kws={'label': 'Spearman Correlation'}, vmin=-0.5,
vmax=0.5, center=0)

    plt.title('Spearman Correlation Matrix', fontsize=16)
    plt.xticks(ticks=np.arange(len(labels)), labels=labels, fontsize=12,
rotation=45, ha='right')
    plt.yticks(ticks=np.arange(len(labels)), labels=labels, fontsize=12,
rotation=0)

    plt.tight_layout(rect=[0, 0.05, 1, 0.95])
    plt.savefig('significant_spearman_correlation_matrix.png', dpi=300)
    plt.show()

corr_df = pd.DataFrame({
    'lang_prod': [1.000000, -0.020490, 0.220389, -0.211314, 0.072364,
0.040644, 0.116463, 0.086243, 0.239944, -0.342745, 0.293728, 0.033088],

```

```

    'esl_level': [-0.020490, 1.000000, -0.119092, -0.040316, -0.375105, -
0.397961, -0.312367, -0.331844, -0.250827, -0.287578, -0.121769, -0.173254],
    'level_of_immersion': [0.220389, -0.119092, 1.000000, 0.272166, 0.108501,
0.061429, 0.140587, 0.054420, 0.089052, 0.067938, 0.177317, 0.116872],
    'modality': [-0.211314, -0.040316, 0.272166, 1.000000, 0.072138,
0.141066, -0.038199, 0.004650, -0.153976, 0.243541, -0.263521, 0.121247],
    'hints': [0.072364, -0.375105, 0.108501, 0.072138, 1.000000, 0.473122,
0.467656, 0.419561, 0.323959, 0.506870, 0.275660, 0.070974],
    'trans_phr': [0.040644, -0.397961, 0.061429, 0.141066, 0.473122,
1.000000, 0.275077, 0.557867, 0.267499, 0.341729, 0.167245, 0.197571],
    'def_word': [0.116463, -0.312367, 0.140587, -0.038199, 0.467656,
0.275077, 1.000000, 0.344203, 0.558358, 0.036679, 0.282073, -0.012884],
    'list_phr': [0.086243, -0.331844, 0.054420, 0.004650, 0.419561, 0.557867,
0.344203, 1.000000, 0.312297, 0.210730, 0.157978, 0.221896],
    'list_word': [0.239944, -0.250827, 0.089052, -0.153976, 0.323959,
0.267499, 0.558358, 0.312297, 1.000000, -0.079544, 0.524011, 0.058714],
    'send_hint': [-0.342745, -0.287578, 0.067938, 0.243541, 0.506870,
0.341729, 0.036679, 0.210730, -0.079544, 1.000000, -0.095833, 0.065619],
    'cor_wr': [0.293728, -0.121769, 0.177317, -0.263521, 0.275660, 0.167245,
0.282073, 0.157978, 0.524011, -0.095833, 1.000000, 0.106256],
    'google': [0.033088, -0.173254, 0.116872, 0.121247, 0.070974, 0.197571, -
0.012884, 0.221896, 0.058714, 0.065619, 0.106256, 1.000000]
}, index=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
        'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
        'Corrective Writing', 'Google Translate'])

p_values_df = pd.DataFrame({
    'lang_prod': [0.000000, 0.820571, 0.009, 0.018000, 0.422564, 0.652687,
0.195866, 0.338912, 0.007035, 0.000091, 0.000885, 0.714131],
    'esl_level': [0.820571, 0.000000, 0.185892, 0.655310, 0.000016, 0.000004,
0.000390, 0.000156, 0.004781, 0.001146, 0.176127, 0.053333],
    'level_of_immersion': [0.009, 0.185892, 0.000000, 0.002137, 0.228415,
0.496159, 0.117866, 0.546664, 0.323351, 0.451567, 0.047901, 0.194289],
    'modality': [0.018000, 0.655310, 0.002137, 0.000000, 0.424018, 0.116605,
0.672341, 0.958954, 0.086447, 0.006203, 0.002985, 0.178001],
    'hints': [0.422564, 0.000016, 0.228415, 0.424018, 0.000000, 0.000000,
0.000000, 0.000000, 0.000228, 0.000000, 0.001861, 0.431552],
    'trans_phr': [0.652687, 0.000004, 0.496159, 0.116605, 0.000000, 0.000000,
0.001904, 0.000000, 0.002563, 0.000096, 0.062290, 0.027207],
    'def_word': [0.195866, 0.000390, 0.117866, 0.672341, 0.000000, 0.001904,
0.000000, 0.000085, 0.000000, 0.684672, 0.001438, 0.886602],
    'list_phr': [0.338912, 0.000156, 0.546664, 0.958954, 0.000000, 0.000000,
0.000085, 0.000000, 0.000392, 0.018328, 0.078482, 0.012881],
    'list_word': [0.007035, 0.004781, 0.323351, 0.086447, 0.000228, 0.002563,
0.000000, 0.000392, 0.000000, 0.377888, 0.000000, 0.515426],
    'send_hint': [0.000091, 0.001146, 0.451567, 0.006203, 0.000000, 0.000096,
0.684672, 0.018328, 0.377888, 0.000000, 0.287725, 0.467191],
    'cor_wr': [0.000885, 0.176127, 0.047901, 0.002985, 0.001861, 0.062290,
0.001438, 0.078482, 0.000000, 0.287725, 0.000000, 0.238243],
    'google': [0.714131, 0.053333, 0.194289, 0.178001, 0.431552, 0.027207,
0.886602, 0.012881, 0.515426, 0.467191, 0.238243, 0.000000]
}, index=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
        'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
        'Corrective Writing', 'Google Translate'])

```

```

        'Corrective Writing', 'Google Translate'])

labels=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
        'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
        'Corrective Writing', 'Google Translate']

visualize_spearman_matrix(corr_df, p_values_df, labels)

```

```

"""This code visualizes Spearman Correlation Matrix showing Spearman
coefficients adjusted by p-values < 0.05"""
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

def visualize_spearman_matrix(corr_df, p_values_df, labels):
    significance_mask = (p_values_df < 0.05) & ((corr_df > 0.1) | (corr_df <
-0.1))
    filtered_corr = corr_df.where(significance_mask, other=0)
    cmap = sns.diverging_palette(220, 20, as_cmap=True)
    mask = np.triu(np.ones_like(filtered_corr, dtype=bool))
    plt.figure(figsize=(12, 10))
    sns.heatmap(filtered_corr, annot=True, mask=mask, cmap=cmap, fmt=".2f",
linewidths=0.5,
                cbar_kws={'label': 'Spearman Correlation'}, vmin=-0.5,
vmax=0.5, center=0)

    plt.title('Spearman Correlation Matrix\nfor p-value < .05', fontsize=16)
    plt.xticks(ticks=np.arange(len(labels)), labels=labels, fontsize=12,
rotation=45, ha='right')
    plt.yticks(ticks=np.arange(len(labels)), labels=labels, fontsize=12,
rotation=0)
    plt.tight_layout(rect=[0, 0.05, 1, 0.95])
    plt.savefig('significant_spearman_correlation_matrix.png', dpi=300)
    plt.show()

corr_df = pd.DataFrame({
    'lang_prod': [1.000000, -0.020490, 0.220389, -0.211314, 0.072364,
0.040644, 0.116463, 0.086243, 0.239944, -0.342745, 0.293728, 0.033088],
    'esl_level': [-0.020490, 1.000000, -0.119092, -0.040316, -0.375105, -
0.397961, -0.312367, -0.331844, -0.250827, -0.287578, -0.121769, -0.173254],
    'level_of_immersion': [0.220389, -0.119092, 1.000000, 0.272166, 0.108501,
0.061429, 0.140587, 0.054420, 0.089052, 0.067938, 0.177317, 0.116872],
    'modality': [-0.211314, -0.040316, 0.272166, 1.000000, 0.072138,
0.141066, -0.038199, 0.004650, -0.153976, 0.243541, -0.263521, 0.121247],
    'hints': [0.072364, -0.375105, 0.108501, 0.072138, 1.000000, 0.473122,
0.467656, 0.419561, 0.323959, 0.506870, 0.275660, 0.070974],
    'trans_phr': [0.040644, -0.397961, 0.061429, 0.141066, 0.473122,
1.000000, 0.275077, 0.557867, 0.267499, 0.341729, 0.167245, 0.197571],
    'def_word': [0.116463, -0.312367, 0.140587, -0.038199, 0.467656,
0.275077, 1.000000, 0.344203, 0.558358, 0.036679, 0.282073, -0.012884],
    'list_phr': [0.086243, -0.331844, 0.054420, 0.004650, 0.419561, 0.557867,
0.344203, 1.000000, 0.312297, 0.210730, 0.157978, 0.221896],

```

```

    'list_word': [0.239944, -0.250827, 0.089052, -0.153976, 0.323959,
0.267499, 0.558358, 0.312297, 1.000000, -0.079544, 0.524011, 0.058714],
    'send_hint': [-0.342745, -0.287578, 0.067938, 0.243541, 0.506870,
0.341729, 0.036679, 0.210730, -0.079544, 1.000000, -0.095833, 0.065619],
    'cor_wr': [0.293728, -0.121769, 0.177317, -0.263521, 0.275660, 0.167245,
0.282073, 0.157978, 0.524011, -0.095833, 1.000000, 0.106256],
    'google': [0.033088, -0.173254, 0.116872, 0.121247, 0.070974, 0.197571, -
0.012884, 0.221896, 0.058714, 0.065619, 0.106256, 1.000000]
}, index=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
'Corrective Writing', 'Google Translate'])

p_values_df = pd.DataFrame({
    'lang_prod': [0.000000, 0.820571, 0.009, 0.018000, 0.422564, 0.652687,
0.195866, 0.338912, 0.007035, 0.000091, 0.000885, 0.714131],
    'esl_level': [0.820571, 0.000000, 0.185892, 0.655310, 0.000016, 0.000004,
0.000390, 0.000156, 0.004781, 0.001146, 0.176127, 0.053333],
    'level_of_immersion': [0.009, 0.185892, 0.000000, 0.002137, 0.228415,
0.496159, 0.117866, 0.546664, 0.323351, 0.451567, 0.047901, 0.194289],
    'modality': [0.018000, 0.655310, 0.002137, 0.000000, 0.424018, 0.116605,
0.672341, 0.958954, 0.086447, 0.006203, 0.002985, 0.178001],
    'hints': [0.422564, 0.000016, 0.228415, 0.424018, 0.000000, 0.000000,
0.000000, 0.000000, 0.000228, 0.000000, 0.001861, 0.431552],
    'trans_phr': [0.652687, 0.000004, 0.496159, 0.116605, 0.000000, 0.000000,
0.001904, 0.000000, 0.002563, 0.000096, 0.062290, 0.027207],
    'def_word': [0.195866, 0.000390, 0.117866, 0.672341, 0.000000, 0.001904,
0.000000, 0.000085, 0.000000, 0.684672, 0.001438, 0.886602],
    'list_phr': [0.338912, 0.000156, 0.546664, 0.958954, 0.000000, 0.000000,
0.000085, 0.000000, 0.000392, 0.018328, 0.078482, 0.012881],
    'list_word': [0.007035, 0.004781, 0.323351, 0.086447, 0.000228, 0.002563,
0.000000, 0.000392, 0.000000, 0.377888, 0.000000, 0.515426],
    'send_hint': [0.000091, 0.001146, 0.451567, 0.006203, 0.000000, 0.000096,
0.684672, 0.018328, 0.377888, 0.000000, 0.287725, 0.467191],
    'cor_wr': [0.000885, 0.176127, 0.047901, 0.002985, 0.001861, 0.062290,
0.001438, 0.078482, 0.000000, 0.287725, 0.000000, 0.238243],
    'google': [0.714131, 0.053333, 0.194289, 0.178001, 0.431552, 0.027207,
0.886602, 0.012881, 0.515426, 0.467191, 0.238243, 0.000000]
}, index=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
'Corrective Writing', 'Google Translate'])

labels=['Language Production', 'ESL Level', 'Immersion(VR vs desktop)',
'Modality(speak vs write)', 'Hints',
'Phrase Translation', 'Word Definition', 'Phrase Pronunciation',
'Word Pronunciation', 'Send Hints',
'Corrective Writing', 'Google Translate']

visualize_spearman_matrix(corr_df, p_values_df, labels)

```