

Immersive Virtual Reality for L2 Pragmatics Instruction
in an Intensive English Program at
a US University
by
Emilia Gracia

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2024 by the
Graduate Supervisory Committee:

Bryan Smith, Chair
Matthew Prior
Marta González-Lloret

ARIZONA STATE UNIVERSITY

May 2024

ABSTRACT

This study examines the impact of High-Immersion Virtual Reality (HiVR) as a tool for teaching English L2 pragmatics to international students enrolled in an Intensive English Program at a US university. This research focuses on integrating HiVR into explicit L2 pragmatics instructional treatments. A quasi-experimental design was employed to assess the pragmatic performance of students in both experimental (HiVR) and control groups across four pragmatics tasks during an eight-week intermediate level course. The findings show a significant advantage for the experimental group, suggesting that HiVR practice enhances learning outcomes in L2 pragmatics. Qualitative feedback revealed participants' enjoyment and perceived realism of VR practice, highlighting the potential of HiVR for simulating real-life conversational scenarios. This study contributes to the field of Computer-Assisted Language Learning (CALL) by demonstrating the efficacy of HiVR in improving L2 pragmatic competence.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER	
1 INTRODUCTION	1
Background.....	1-2
Problem.....	2-3
Solution.....	3-4
Purpose of Study.....	4-5
Overview.....	5
2 LITERATURE REVIEW	2
Theoretical Framework for L2 Pragmatics.....	6-11
L2 Pragmatics in Study Abroad Contexts.....	12-14
L2 Pragmatics in Computer-Assisted Language Learning.....	14-20
Virtual Reality (VR).....	20-21
Research Questions.....	29-30
3 METHOD	3
Research Context.....	30-33
Description of Instructional Treatment Material.....	33-36
Rationale for Instructional Treatment.....	40-42
Research Instruments.....	42 -46
Data Collection Timeline.....	46-47

CHAPTER	Page
Data Collection Procedures.....	47-48
Data Analysis.....	48-51
Chapter Summary.....	51
4 RESULTS	4
Quantitative Analysis.....	52-54
Research Question 1.....	54-59
Research Question 2.....	60-61
Qualitative Analysis.....	61-66
Summary of Quantitative Results.....	66-67
Summary of Qualitative Results.....	67
5 DISCUSSION AND CONCLUSION.....	5
Research Question One.....	68-69
Research Question Two.....	69-71
Research Question Three.....	71-73
Pedagogical Implications.....	73-76
Limitations.....	76-77
Method of Evaluation.....	77-80
Recommendations for Future Research.....	80-81
Conclusions.....	81
REFERENCES	6

APPENDIX

A. IRB APPROVAL 93-94

B. ROLE-PLAY TESTS 95-97

C. RATING SCALES..... 98-99

D. SURVEY QUESTIONS..... 100-101

E. RESPONSES TO SURVEY.....102-108

F. PRACTICE ROLE-PLAY SCRIPTS..... 109-113

LIST OF TABLES

Table		Page
1.	Participant Demographic Information	33
2.	Timeline	39
3.	Interrater Agreement for Appropriateness Role-Play Tests.....	53
4.	Interrater Agreement for Grammar Role-Play Tests.....	54
5.	Pre and Posttest Descriptive Statistics for Appropriateness.....	55
6.	Pre and Posttest Descriptive Statistics for Grammar.....	56
7.	Wilcoxon Signed-Ranks Tests for Appropriateness in Experimental and Control Groups.....	57
8.	Wilcoxon Signed-Ranks Tests for Grammar in Experimental and Control Groups.....	59
9.	Mann-Whitney U Test for Role-play Test Gains Scores for Appropriateness	60
10.	Mann Whitney U Test Results for Role-play Test Gains Scores for Grammar	61
11.	Survey Items with Likert Scale Ratings (Questions 1-4)	62
12.	Distribution of Multiple-Choice Responses (Question 5)	63
13.	Distribution of Multiple-Choice Responses (Question 6)	63

LIST OF FIGURES

Figure	Page
1. Example of a Role-Play Script in ImmerseMe	36
2. Example of Pronunciation Activity in ImmerseMe	36
3. Frequency of Coded Themes for Short-answer Question 7.....	64
4. Frequency of Coded Themes for Short-answer Question 8.....	65
5. Frequency of Coded Themes for Short-answer Question 9.....	66

CHAPTER 1

INTRODUCTION

1.1 Background

In second/foreign language curriculums and classrooms, much attention is given to aspects of language such as grammar, reading, writing, listening, and speaking, as they all form part of one's linguistic repertoire and are necessary for communication in any language. However, there is one highly necessary language skill that has been traditionally excluded from foreign language textbooks and instruction, and that is the skill of second language (L2) pragmatics (González-Lloret, 2021). L2 pragmatics, according to Bardovi-Harlig (2013) is “knowing how to say what to whom when” (p. 68). Speakers of any language require communicative competence in order to communicate effectively in their contexts, whether the language they are speaking be their first, second, third, etc. (Hymes, 1972). The idea that communicative competence is necessary for successful language use in a context where learners interact with L1 users is widely accepted in the field of Applied Linguistics (Hymes, 1972; Canale & Swain, 1980; Bachman & Palmer, 1996) and pragmatic competence forms part of communicative competence.

Leech (1983) and Thomas (1983) theorize that pragmatic knowledge comprises two types of linguistic resources: pragmalinguistics (use of linguistic functions) and sociopragmatics (assessment of context in which the language is used). L2 pragmatic competence is a learner's ability to use linguistic forms in the target language in a way that is contextually and socially appropriate (Ekiert et al., 2018; Kasper & Rose, 2002).

Lack of pragmatic competence can result in pragmatic failure, leading to miscommunication with potentially negative social consequences (Kasper & Rose, 2002). Therefore, learners need to understand and learn to use the pragmatics of the L2 in order to become competent communicators in the target language, which is why it should be taught in foreign and second language classrooms.

1.2 Problem

L2 learners in Study Abroad (SA) contexts are contextually situated to develop pragmatic competence in the target language, being as they are afforded opportunities to communicate with speakers of the target language in a context where the target language is spoken. Though these communicative opportunities exist, and much more so in a SA context than in a Foreign Language (FL) context, this does not mean that learners implicitly acquire L2 pragmatic knowledge in the wild as a mere consequence of being physically situated in a target language context (Vidal & Shively, 2019; DuFon, 1999). The context of the current study is a SA context (from the perspective of participants of this study), specifically, an Intensive English Program (IEP) at a US university, where learners are studying English for admission to the university. Learners in this context receive around 21 hours of intensive English instruction per week in eight-week sessions, where they receive instruction in academic English for reading, writing, listening and speaking. Most learners in the IEP plan to complete an undergraduate degree in the US, therefore they need a certain level of English proficiency in order to achieve their goal.

Having served as a full-time instructor in this IEP since 2013, I have observed that the four-skills courses prepare students quite well for academic success, however,

competence in academic English does not necessarily transfer to competent use of English in social contexts. Over the years, my students have occasionally commented on how difficult it is for them to make friends with local, native English speaker students, especially those who are native English speakers. They have reported not knowing what to say or how to express themselves without potentially offending someone or that they simply lack confidence in using English in social contexts. As is the case in most educational settings, L2 pragmatics is not specifically taught in this IEP context, and learners in the context of this study need pragmatic knowledge in order to be socially successful.

1.3 Solution

Learners in the context of the current study have one elective course dedicated to communication, and it is focused on conversational exchanges in English, however, L2 pragmatics are not taught explicitly in this course. Nonetheless, there is much potential for L2 pragmatics integration into the course curriculum. The question of how to teach L2 pragmatics and what methods and materials to use needs to be answered before planning an instructional intervention. L2 pragmatics have been considered by many as difficult to teach due to lack of instructional materials or access to speakers of the target language, however, the field of Computer-Assisted Language Learning (CALL) has much to offer in terms of solutions to this problem. Digital tools afford learners the opportunity to interact with speakers of the target language without physical travel, in addition to infinite practice and feedback features. Several CALL-based studies have been done to test the effectiveness of the affordances of different digital tools for L2

pragmatics learning. Email, instant messaging, digital games, Virtual Reality (VR) and computer simulations have been tested in studies such as those of Sydorenko et al. (2017), Sykes (2009), Tang and Taguchi (2020), and García-Gómez (2020), to mention a few, however, to date, no published research for testing the effectiveness of high immersion Virtual Reality for explicit L2 pragmatics instruction currently exists.

1.4 Purpose of Current Study

L2 pragmatics research in the field of CALL exists, but it remains scarce. In addition, research concerning L2 pragmatics with VR specifically is even more scarce. Existing research includes a study exploring the usability of immersive VR to evaluate pragmatic competence in request making (Taguchi, 2021), a study that compares dispreferred speech acts in computer-based and VR environments of native and nonnative speakers (Taguchi, 2022a), and a study that explores the usability of immersive VR for the development of pragmatic tasks (Taguchi, 2022b). These studies are largely exploratory and none of them test the usability of immersive VR, in a classroom context, as the practice component of explicit L2 pragmatics instruction. Therefore, the purpose of the current study is to fill this gap by contributing to a body of knowledge from which learners, practitioners, and researchers can draw insights regarding the classroom implementation of immersive VR. This dissertation study examines the effectiveness of integrating VR into explicit pragmatics instructional treatment as a tool for practicing speech act sequences (refusals, requests, and apologies) in the form of role-plays in a fully immersive environment where learners get repeated opportunities for practice and feedback. The aim of the current study is to compare the learning effects of participants

who used VR to practice pragmatic role-play tasks to the effects of participants who practiced the same role-play tasks in a traditional classroom environment. Both experimental and control groups receive the same explicit instructional pragmatics treatment; the experimental group completes their practice activities in a fully immersive VR environment while the control group completes the same practice activity in a traditional classroom setting with peers. This quasi-experimental, instructional intervention takes place over four treatment periods, during an eight-week session, in the context described in section 1.2. Another aim of this study is to see if learners improve their pragmatic performance in face-to-face conversations with domestic native speaker students at their same university as a result of having practiced in VR. In order to measure the effectiveness of VR use for this purpose, participants' pragmatic performance is measured by means of role-play tests which they perform with a domestic native speaker student partner. The results of this study inform future integration of VR into classroom environments and contribute to a sorely lacking area of research.

1.4 Overview

This dissertation is organized into five chapters. This chapter was the first of five and its purpose was to provide background knowledge about the context of the study and L2 pragmatics instruction in CALL. Chapter 2 is a review of the current literature, where I discuss the theoretical framework of the study and existing literature in L2 pragmatics instruction in study abroad contexts and CALL, followed by studies that used VR for language learning purposes and the research questions for this dissertation study. Chapter 3 discusses the methods used to carry out the current study. Research processes and

research procedures are explained in detail. Chapter 4 presents the results of the statistical analysis for the data collected for this study, answering the research questions, and Chapter 5 discusses these findings in detail, in addition to pedagogical implications and limitations of the study.

CHAPTER 2

LITERATURE REVIEW

This chapter presents a review of the literature pertaining to the theoretical basis of this study and previous research in L2 pragmatics instruction in CALL and VRALL. First, I define L2 pragmatics and discuss the theoretical frameworks for the current study (section 2.1). Next, I provide an overview of existing literature regarding L2 pragmatics instruction in Study Abroad (SA) contexts (section 2.3), followed by a review of L2 pragmatics research in CALL (section 2.4). Last, I will give an overview of existing CALL research in which VR is utilized (known as VRALL, or Virtual Reality Assisted Language Learning, section 2.5) and state my research questions for the current study (section 2.6).

2.1.1 Theoretical Framework for L2 Pragmatics

L2 or interlanguage pragmatics, a subfield of Second Language Acquisition (SLA) primarily concerned with how language is used in context, has evolved over time and inspired an array of research studies. Grounded in Hymes (1962) theoretical framework for studying communication as a phenomenon that occurs in real-life contexts bound to social rules and roles, some of the earliest theoretical frameworks for L2 pragmatics include the Speech Act Theory (Austin, 1962; Searle, 1976), Grice's maxims (Grice, 1975) Politeness Theory (Brown & Levinson, 1987). Of the aforementioned theories, one of the most influential in early L2 pragmatics research was The Speech Act Theory, which conceptualizes speaker utterances as illocutionary acts that are categorized as either requests, refusals, apologies, or promises, (among other concepts such as felicity

conditions and performatives vs. constatives), (Searle, 1969, 1976). L2 pragmatics studies that emerged in the 1970's and 1980's was influenced by these theories and they were concerned with learner pragmatic transfer (from L1 to L2), and the L2 users' speech act production in comparison to that of native speakers. This L2 pragmatics research (conducted before the 1990s) was reflective of the Cognitive-Philosophical view, one of two perspectives of pragmatics.

The concept of pragmatic knowledge prevailed in Leech (1983) and Thomas (1983) and as a construct of two aspects: pragmalinguistics (the language forms needed to communicate) and sociopragmatics (the social conventions and context in which language is used). Similarly, the concept of communicative competence emerged in the works of Canale and Swain (1980) and Bachman and Palmer (1996) which emphasized the idea that being a competent communicator requires more than just knowledge of linguistic forms, it requires one to use language appropriately within a specific sociocultural context. In short, one must have pragmatic knowledge as defined by Thomas (1983) and Leech (1983) in order to communicate competently as pragmatic competence is essential to communicative competence (Canale & Swain, 1980; Bachman & Palmer, 1996). Following this, L2 pragmatic competence became more widely studied in the 1990s and 2000s as researchers began measuring it with data collection instruments such as Discourse Completion Tests (DCTs), role-plays and verbal reports.

More recently, the concept of interactional competence has been added to the definition of pragmatic competence. Being pragmatically competent now includes one's ability to be flexible and adapt their choice in language use according to the context and

interlocutors. Not only should a pragmatically competent communicator know which linguistic forms and speech acts to use, they should also have interactional skills such as turn-taking, repair strategies and management of topic, as these are necessary for effective communication (Young, 2011). With this current and integrated concept of pragmatic competence, data collection methods such as Conversation Analysis (CA) lend themselves to L2 pragmatics research that analyzes authentic communication in the wild (Kasper, 2006).

Finally, the concept of L2 pragmatic competence that I have chosen to adopt for this dissertation study (also mentioned in Chapter 1) is a popular and simple from Bardovi-Harlig (2013), stating that pragmatic competence is “knowing how to say what to whom when,” (p. 68). This definition acknowledges both the pragmalinguistic and sociopragmatic dimensions theorized by Leech (1983) and Thomas (1983), and it encompasses both the Cognitive-Philosophical and Sociocultural-Interactional perspectives of L2 pragmatics. Likewise, interactional competence (Young, 2011) forms an integrated part of my understanding of pragmatic competence, and thus it is reflected in my data collection instruments utilized in the current study. Now that I have provided background information and a clear definition of L2 pragmatics and pragmatic competence, I will give an overview of studies concerning instructed L2 pragmatics in the following section.

2.1.2 Theoretical basis for L2 pragmatics instruction

Reasons for teaching L2 pragmatics in classroom settings do not differ from reasons for teaching any other aspect of a second or foreign language. In fact, several

SLA theories commonly applied to the teaching and learning of other linguistic features also apply to the teaching and learning of L2 pragmatics, especially explicit instruction. To date, there exists a consensus among scholars in the field that L2 pragmatics is teachable and that learners who receive instruction show higher learning outcomes than those who do not (Kasper & Rose, 1999), and that explicit instruction has been shown (in the majority of cases) to have greater learning outcomes than implicit instruction according to several meta-analysis studies (Plonsky & Zhuang, 2019; Taguchi, 2015; Taguchi, 2011; Takahashi, 2010b) and reviews (Jeon & Kaya, 2006; Rose, 2005). In Taguchi's (2015) review of studies on instructed L2 pragmatics, they compared teaching methods (and different ways to teach explicitly) and the findings indicate that learners benefit the most from explicit instruction that includes metapragmatic information and form-focused practice activities where they have opportunities to produce and interact (i.e., role-plays). This description of explicit instruction fits Kasper's (2001) construct (specific to L2 pragmatics) and follows the same pattern as Ellis' (2005) explicit form-focused instruction approach (which is applied to various language skills).

The type of explicit instruction described above is well supported by longstanding SLA theories, such as Schmidt's (1990) noticing hypothesis, Swain's (1995) output hypothesis, Ellis' (2008) variability theory, Long's (1981) interactionist approach, and Krashen's (1982) input hypothesis. Metapragmatic information, the first aspect of explicit L2 pragmatics instruction described in the previous paragraph, is supported by Schmidt's (1990) noticing hypothesis, which claims that in order for learners to take in any given input, they must consciously notice it. Metapragmatic explanation and awareness-raising

activities do this specifically. For example, the instructor presents information about a form and explains why and how the form is used so that learners are aware of it.

The second aspect of explicit instruction is communicative activities that require learners to produce the target form. This aspect is well supported by Swain's (1995) output hypothesis (which is related to noticing as well) states that learners acquire forms when they produce them, and that production of language triggers noticing of forms; essentially, production is necessary for learning. Likewise, Ellis' (2008) variability theory supports the productive practice aspect of explicit instruction; Ellis (2008) claims that learner output varies during language acquisition when experimenting with different structures. Practicing role-plays in class, especially open role-plays, gives learners the opportunity to experiment with the form they learned and observe how the interlocutor responds to them.

Next, Long's (1981) interactionist approach could be applied here as foreign language courses are typically organized to maximize learner-to-learner interaction. Likewise, students in SA contexts often learn alongside students of other L1s, and must negotiate meaning in the target language. Last, a specific aspect of Krashen's (1982) input hypothesis supports the productive practice of forms, especially in the context of explicit L2 pragmatics instruction. When learners are given the opportunity to practice producing target pragmatic forms in class, simulating a real-life conversation, their affective filter will likely be much lower when doing so in class, among peers, than in real-life scenarios. In this case, learners get the opportunity to learn in a low-stress environment where they can practice without unforeseen negative social consequences.

Thus far, it has been established through various theoretical frameworks that pragmatic competence is essential to communicative competence, making it important to learn if one wishes to become proficient in any language. In addition, I have discussed research showing that L2 pragmatics are teachable, learners benefit from instruction, specifically explicit, and that this type of instruction is well-supported by SLA theories. In spite of these observations, L2 pragmatics is rarely (if ever) taught in foreign language classrooms or study abroad programs, and does not appear in mainstream textbooks (Bardovi-Harlig, 2017; González-Lloret, 2021). Reasons for this vary, however, it is likely that pragmatics are considered difficult to teach, foreign language contexts lack access to authentic input, or it may be assumed that pragmatics are acquired naturally outside of the classroom (which is an assumption often made in SA contexts, see section 2.2.1). Also, it is difficult for textbook editors and curriculum designers to reach an agreement about *whose* pragmatics should be used as a model, as L2 pragmatics vary across regions and sociocultural contexts (Bardovi-Harlig, 2001). Certainly, it is difficult to learn L2 speech act sequences (making requests, apologizing, refusing etc.), politeness, implicatures (and other pragmatic functions) because interaction is fluid, dynamic, and dependent on interlocutors, and there is much variability within different contexts. Nonetheless, even if learners do have access to speakers of the target language and are living in the context where the target language is spoken, there is no guarantee that they will learn pragmatics (Vidal & Shively, 2019). For this reason, I have given an overview of L2 pragmatics learning in SA contexts in the following section.

2.3 L2 Pragmatics in Study Abroad Contexts

Indeed, it is worth noting that Study Abroad (SA) contexts, pose significant opportunities for authentic interaction and observation of pragmatic norms of the target L2, and this may lead some to believe that there is no need for pragmatics instruction in a SA context. Nonetheless, it cannot be assumed that learners will simply *notice* what language is pragmatically appropriate and be able to use it immediately and effectively as a result of implicit socialization - in order for noticing (Schmidt, 1990) to result in raised awareness and knowledge, learners need explicit instruction of some kind. For example, in Shively (2011), a study that examined L2 Spanish language pragmatics for service encounters in a SA context, it was observed that not all settings permitted observation or the opportunity to notice local language user pragmatics. For example, learners could not listen to conversations at the bank between local customers and bankers as this tends to be a more private interaction (unlike customer service interactions at the supermarket). Therefore, opportunities to imitate native speaker pragmatics are not as abundant as one might imagine.

Likewise, it cannot be assumed that all learners in SA contexts will have the same experiences and opportunities. Study Abroad language exposure varies in many aspects, such as living accommodations, language instruction, previously acquired knowledge and experiences in the target language, individual motivation, frequency of communication with speakers of their L1, and social networks of other international students (Vidal & Shively, 2019). For example, some SA students live with a host family while others may choose to live in the school's dormitories or share an off-campus living space with roommates. Also, language programs differ in their instructional methods and required

class time. SA learners may or may not attend classes in the target language with domestic students, which limits exposure to authentic pragmatics input.

Next, rich exposure to the L2 cannot guarantee pragmatic competence due to the fact that learners, as individuals of various ideologies and backgrounds, vary in personality, gender, intellectual abilities, religious beliefs, motivation, self-perception, and position within the sociocultural context in which they are studying; these factors influence the interactional exposure and practice experienced by learners (DuFon, 1999; Kinginger, 2009). In addition, it is possible that even if learners do know the pragmatic rules of the target language and culture, they may refuse to use them if they do not identify with said rules and possess contradicting beliefs about them (as seen in Ishihara & Tarone, 2009; Ishihara, 2019; Siegal, 1996). Although opportunities for interaction with native speakers of the target language do provide a source of authentic input, target language users or native speakers may not offer corrective feedback (Hassall, 2013), or they may perceive L2 learners as not capable of understanding pragmatically complex language and as a result speak to L2 learners in a simplified form of language (or “teacher talk”) that is not representative of the target language pragmatics as seen in Iino (2006). Therefore, even if a learner in a study abroad context receives many opportunities to interact socially and acquire pragmatic knowledge, awareness, and competence, there is no guarantee that this will happen solely based on the fact that they are physically positioned in a SA abroad context.

Being as the SA context alone is not sufficient for acquiring target language pragmatic competence, the possibility of L2 pragmatics instruction in SA contexts is

worth considering as it has been shown to be effective even in non-SA classroom contexts, as seen in Plonsky and Zhuang's (2019) meta-analysis study. Studies such as those of Alcón-Soler (2015), Winke and Teng (2010), Halenko and Jones (2011), and Bouton (1999) compared instructed groups to non-instructed groups in SA contexts and show positive learning effects for the instructed groups. Though these studies are few in comparison to those focused-on learning of L2 pragmatics in the wild in SA contexts (e.g. Shively, 2011), they serve as a starting point and call for further research in L2 pragmatics instruction in SA contexts, the context in which this dissertation study has been situated.

2.4 L2 Pragmatics in Computer-Assisted Language Learning (CALL)

Thus far, I have identified the need for L2 pragmatics instruction in foreign or second language learning and specifically in a SA context. Next, when considering ways in which L2 pragmatics can be learned, the many possibilities and affordances of technological tools for foreign language pragmatics instruction cannot go unnoticed. For this reason, I provide an overview of the existing literature on L2 pragmatics in CALL.

Certain digital technologies offer solutions to the instructional challenges of L2 pragmatics, as the internet allows us carry out telecollaboration projects across continents, see the world through social networking applications, experience the three Ps of culture (products, practices, perspectives) in digital spaces (Taguchi & Roever, 2017). It is now possible to learn sociocultural norms of the target language culture and interact via Computer-Mediated Communication (CMC) and Synchronous Computer-Mediated Communication (SCMC) with L1 speakers of the target language without traveling,

making the present era a prime-time frame to conduct CALL research in L2 pragmatics. However, according to González-Lloret (2019) studies for pragmatics in CALL are very few, and reports that “since 2016, only 7% of the articles published in *Language Learning & Technology* focused on pragmatics, and the percentage is even lower for *CALICO Journal* (2.2%) and *ReCALL* (1.8%)” (p. 113).

From the existing literature, it is known that many CALL-based studies show positive effects for L2 pragmatics learning using CMC and SCMC tools in different contexts and environments (Akiyama & Cunningham, 2017). Studies showing positive learning effects with email being the CMC tool of choice for L2 pragmatics research include those of Chang and Hsu (1998), Chen (2006), Biesenbach-Lucas (2006, 2007), Alcón-Soler (2018), Félix-Brasdefer (2012), Chen (2015), Nguyen (2018), Usó-Juan (2022), Codina-Espurz and Salazar-Campillo (2019), Salazar Campillo (2018), Salazar-Campillo and Codina-Espurz (2018), and Winans (2020). It is likely that email has been used so frequently as a CMC tool in L2 pragmatics CALL research due to its ubiquity and ease of use in data collection.

With email appearing to be the most frequently used CMC digital tool in L2 pragmatics research in CALL, the most studied pragmatics topic of interest in the aforementioned studies appears to be the speech act sequence of making requests and politeness in making requests (Chang and Hsu, 1998; Biesenbach-Lucas, 2006; Alcón-Soler, 2018; Félix-Brasdefer, 2012; Chen, 2015; Dombi, 2020; Nguyen, 2018; Usó-Juan, 2022; Winans, 2020; Biesenbach-Lucas, 2007; Chen, 2006), while some focused on openings and closings (Codina-Espurz & Salazar-Campillo, 2019; Salazar Campillo,

2018; Salazar-Campillo & Codina-Espurz, 2018). Making requests via email is a pragmatic task that is useful and necessary for learners in many contexts, especially university students who are studying abroad. The practicality and ubiquity of this particular task has made it a research topic of interest. For learners of English, writing a polite email making a request to someone in a higher position of power can be difficult due to the expected indirectness and inconsistency in levels of formality from interlocutors. For example, if a student at a US university writes an email to their professor (making a request) using language that is considered polite and formal, the professor may respond with informal language (i.e. referring to themselves by their first name when closing the email, being direct and brief in their communication, etc.) and this can be confusing for L2 learners because they cannot be sure if they are expected to continue the communication exchange with an equally informal response. Examples such as this are reasons why learners need L2 pragmatics instruction; explicit instruction and practice carrying out pragmatic tasks (such as making requests via email) can help learners improve their pragmatic production (Chen, 2015; Alcón-Soler, 2018; Nguyen, 2018).

Other studies concerning directness and indirectness in email communication are those of Dombi (2020) and Winans (2020). Dombi (2020) found that L2 users preferred direct strategies, however, Winans (2020) showed that non-native speakers were rated as being more polite than native speakers when rated by highly-proficient English speakers from the US, Korea, India, and China. This calls for more research in variational pragmatics in CMC contexts. Winans (2020) argues that English L2 pragmatics

instruction ought to be more inclusive of varieties of English as politeness norms change from variety to variety (i.e. American English, Indian English, Australian English, etc.).

Other CALL studies have examined L2 pragmatics learning SCMC environments that use text-based chat/instant messaging have also shown positive learning effects for L2 pragmatics and are fewer in number (than those of asynchronous CMC) but vary in topic of focus. Topics of study include speech act sequences such as refusals and/or apologies (Sykes, 2005; Sykes, 2009; Vilar-Beltrán & Melchor-Couto, 2013), formulaic expressions (Tang & Taguchi, 2020), addressivity (Belz & Kinginger, 2002; Belz & Kinginger, 2003; González-Lloret, 2008; Williams & van Compernelle, 2009; Van Compernelle et al., 2009), politeness (Gonzales, 2013), appropriateness (Cunningham, 2017), and a mixed variety of speech act sequences (Cunningham, 2016; Soares Palmer, 2010; Sykes, 2012). One study that did not show positive learning effects for L2 pragmatics was that of García-Gómez (2020), where English and Spanish speakers interacted on Whatsapp and misunderstanding and overuse of slang led to hostile interactions, which negatively affected L2 pragmatics learning. This finding is insightful in that it reminds researchers and practitioners of the importance of designing a learning environment that is well-structured and includes metapragmatic explanation and well-established rules and guidelines for communication.

Some studies feature gamified elements and Synthetic Immersive Environments (SIEs) (i.e. Sykes, 2009) such as Holden and Sykes' (2012) game *Mentira*, Tang and Taguchi's (2020) *Questaurant*, or Soares Palmer's (2010) use of an already existing game *World of Warcraft*, and Vilar-Beltrán & Melchor-Couto's (2013) and Peterson's

(2018) use of the SIE *Second Life*. These studies showed positive gains in L2 pragmatics learning, but not all showed improvement in pragmatic production. For example, in Sykes (2009) learners demonstrated gains in knowledge of apologies and refusals, but did not show improvement in their ability to produce these speech act sequences. Likewise, some studies showed more than just positive learning effects; Tang and Taguchi's (2021) study not only showed positive effects for learning formulaic expressions in Chinese in a gamified SIE, it also showed greater learning effects for explicit L2 pragmatics instruction. The SIEs in which these SCMC studies took place offered a learning environment where learners could experiment with and develop L2 pragmatics without risking the potential negative social consequences that could result from doing so in face-to-face environments in the wild. Given this unique affordance, it is only logical that more studies for L2 pragmatics learning in SCMC environments be done in the future, especially for classroom contexts where teachers would like to teach L2 pragmatics more effectively and interactively.

Another tool that has been utilized for L2 pragmatics research and teaching is that of computer simulations, especially for oral practice. Computer simulations use Spoken Dialogue Systems (SDS) which permits users to carry out an interactive dialogue with speech characters. Some SDS use Automatic Speech Recognition (ASR) software and Natural Language Processing (NLP) algorithms, allowing users to receive audio and visual input, and produce oral output. The benefits of this type of tool for learning L2 pragmatics is that learners are afforded limitless practice and feedback, and can notice gaps in their production with extended discourse practice. Typically, in computer

simulations, users watch a pre-recorded video, respond orally via audio recording to what they saw, and based on their response, they are taken to the next scenario. There have been different variations of this used in L2 pragmatics research that are less technologically advanced (exclude ASR), such as Sydorenko et al. (2017) and Sydorenko et al. (2020) where after viewing a video user record themselves and then answer a multiple-choice question asking them to indicate what type of response they gave, and based on their selection, a new screen appears with the next scenario.

Qualitative findings from studies using computer simulations for L2 pragmatics learning are positive in that they report the simulations as being useful in their contexts (Sydorenko et al., 2020), authentic, and gave them increased confidence (Sydorenko et al. 2018). One particular finding of interest in Sydorenko et al. (2018) is that learners reported feeling nervous as if they were talking to a real person even though they were not. Quantitative findings from these studies have implications for explicit versus implicit instruction, and directness in L2 requests. Sydorenko et al. (2020) found very little difference in implicit and explicit instruction groups, indicating that learners can learn L2 pragmatics using computer simulation implicitly, but that learners have different needs and some of them need more explicit instruction than others.

As for directness and indirectness in requests, Laughlin et al. 's (2017) study used a computer simulation to observe requests in L1 and L2 users of English, and found that L1 users preferred indirect requests (except when requesting a meeting), and L2 users preferred direct requests. Sydorenko (2015) compared learning requests in computer-delivered structured tasks (CASTS) to learner-to-learner role-plays. They found that

CASTS helped learners to produce forms closer to the model form, but that role-plays rendered more creative language use and helped learners have metapragmatic discussions. Sydorenko et al. (2018) also looked at requests in computer simulated environments, which had positive findings similar to Sydorenko (2015) for learners being able to produce the target form more accurately due to the computer simulation.

Computer simulations have shown, thus far, to have positive learning effects for L2 pragmatics, as they give learners opportunities for practice and feedback, and a learning experience that simulates real-life communication. However, L2 pragmatics in CALL research continues to evolve as new tools with more advanced capabilities become available. A more technologically more evolved form of computer simulations is Virtual Reality (VR), especially immersive VR. Digital technologies such as immersive Virtual Reality (VR) offer great potential for L2 pragmatics learning, and research on this particular use of VR is currently scarce, including only a few studies mentioned earlier (Taguchi, 2021; Taguchi, 2022a; Taguchi, 2022b), hence the need for this dissertation study. In the following section I will give an overview of what is known of VR for language skills other than L2 pragmatics.

2.4 Virtual Reality (VR)

According to Merriam-Webster online dictionary, Virtual Reality is “an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment.” Upon mentioning VR, perhaps what first comes to mind is gaming, however, VR has been widely used for purposes other than entertainment. For example,

VR environments have been used for learning focused on aviation, medical training, business, and education. According to Qiu et al. (2023), many studies using VR for EFL learning have emerged since 2015, and this is not surprising given VR's many affordances: realistic environment, continuous practice and feedback, individualized learning, among others. However, not all VR is the same, so this should be taken into consideration when looking at the different types of VR that have been used in previous studies, as they vary in affordances according to type. Definitions of VR types vary slightly; Qiu et al. (2023) mention three types of VR in their meta-analysis of studies that use VR: (1) non-immersive, (2) semi-immersive, and (3) immersive. Non-immersive refers to desktop versions of VR, semi-immersive refers to Cave Automatic Virtual Environments (CAVE) technology. These two types of VR are also referred to as low-immersion VR or LiVR (Dhimolea et al., 2022; Kaplan-Rakowski & Gruber, 2021).

Finally, immersive VR typically uses head-mounted devices (HMDs) with audio input and output, 3D visual displays, and hand-held devices to give users a fully immersive experience. This type of VR is also known as high-immersion VR or HiVR (Dhimolea et al., 2022). Some VR, such as World of Warcraft, are game-based while others are socially-based. Both types of VR have been used to study foreign language learning and thus are included in this review. It is important to distinguish between these different types of VR when reviewing VR studies for CALL research or VRALL (Virtual-Reality Assisted Language Learning) because HiVR affords learners the opportunity to have an embodied experience, closely imitating real-world interactions that provide authentic communication practice that is context-specific (Kaplan-Rakowski

& Gruber, 2021), while LiVR cannot provide learners with the same level presence and embodiment.

Contrary to HiVR, LiVR is cost effective and readily available making it accessible to several populations of learners and researchers. Nonetheless, as HiVR equipment and software become more prevalent, prices are likely to decrease, and this could result in wider adaptation and implementation of HiVR in foreign language classrooms and VRALL research. Given the many affordances of VR technologies for foreign language learning (FLL) and its ability to provide users with embodied, emotional, close-to-real-life experiences via immersive headset gear (Gall et al., 2021), it is likely that there will be more HiVR studies in the future. As there are not as many VRALL studies with HiVR as there are with LiVR, I used HiVR in this dissertation study in order to address this gap. The following is a review literature of studies using both LiVR and HiVR.

2.5 Virtual Reality Assisted Language Learning (VRALL)

VR research in CALL contexts has been ongoing since the early 2000's (Lan, 2020), and researchers in the field have identified VR for its numerous capabilities in providing learners with simulated experiences they would not otherwise be able to access. Indeed, comparatively little research has been conducted in VR use for CALL purposes, however, the work that has been done was examined in a recent meta-analysis (Qiu et al., 2023) and two systematic reviews concerning this topic (Dhimolea et al., 2022; Parmaxi, 2020). Each has a slightly different focus, but there are various commonalities in their search criteria and results.

Similarly, Qiu et al. (2023), Dhimolea et al. (2022), and Parmaxi (2020) reviewed VR in CALL studies published in 2015 or later, with Qiu et al. (2023) accounting for the vastest range of 2015-2021, while Dhimolea et al. (2022) accounted for studies between 2015-2020, and Parmaxi (2020) had the shortest range of studies published between 2015-2018. All articles feature research based on the specific criteria that the studies must be empirical and published in an academic venue. The authors of the three articles mentioned above concur that there were very few studies with VR in CALL published before 2015, and therefore establish this year as the starting point. At present, these ranges are also indicative of the immaturity of this field. The authors of these also establish that VR for educational purposes is a blossoming field of interest for many researchers and educators in a variety of disciplines (i.e. medicine, aviation, natural sciences, etc.), and that the use of VR for language learning has great potential to afford foreign language learners the opportunity to acquire the target language in an interactive environment. In addition, authors agree that the practice and feedback characteristics that VR affords learners reaches beyond the bounds of what is possible in a traditional classroom environment, therefore hypothesizing that foreign language learning would improve with VR use.

Nonetheless, Qiu et al. (2023), Dhimolea et al. (2022), and Parmaxi (2020) differed in their foci and findings Parmaxi (2020) and Dhimolea et al. (2022) reviewed studies treating a variety of foreign languages, while Qiu et al. (2023) analyzed studies that measured the effects of VR on EFL learning against traditional EFL instruction, including studies with at least one experimental and control group. Though this

differentiates Qiu et al. (2023) meta-analysis from the two systematic reviews, the presence of English is strongly noted in Dhimolea et al. (2022) and Parmaxi (2020), as English was the most studied foreign language as well as the most common native language of participants learning other languages. Another commonality shared in these reviews and meta-analysis was the age of the most studied learner population: college students (ages 18-25). Educational settings for the studies reviewed in these meta-analyses showed Asia as being the most common regional setting, with North America being the second.

Similarities and differences in foci are also apparent in the above-mentioned publications. Dhimolea et al.'s (2022) focus differed from that of Parmaxi (2020) and Qiu et al. (2023) in that it included only high-immersion VR, while the others included non-immersive VR and semi-immersive VR (in addition to high-immersion or full immersion) in their corpus of studies. Parmaxi (2020) and Dhimolea et al. (2023) generally sought to identify VR technologies used, learner populations and settings, target language domains, learner perceptions, benefits and limitations of VR technology use as well as implications for future research and pedagogy. Qiu et al.'s (2023) meta-analysis was slightly narrower, more specific, and less exploratory than the systematic reviews, as it had six research questions which sought to answer the larger question of how effective VR is on EFL instruction and the reasons contributing to effect sizes. Studies that did not include quantitative data were excluded from Qiu et al.'s (2023) meta-analysis, while Parmaxi (2020) and Dhimolea et al. (2022) included qualitative, quantitative and mixed methods studies. The sample sizes of all three reviews were all

relatively small, ranging from 14 -34 studies, and all yielded valuable findings for educators and researchers in the field.

Qualitatively, for learner attitudes and perspectives on VR, Dhimolea et al.'s (2022) review reported a trend in studies with positive findings for learners having an optimistic perception of VR, high motivation, reduction in anxiety, and overall enhanced foreign language learning enjoyment. However, it should be noted that not all VR-based studies show the positive trends mentioned above. For example, Taguchi (2021) found that learners were more anxious and nervous in the VR environment due to the realistic nature of the simulation. Parmaxi (2020) found that in addition to enhanced learning, VR-based studies showed that learners acquired enhanced 21st century skills such as teamwork, collaboration, and problem-solving. Qiu et al. 's (2023) was focused more on learning outcomes of studies (especially EFL studies) and did not report trends in qualitative findings.

One unique and insightful finding from Dhimolea et al. (2022) is that learners must receive VR treatment multiple times in order to show results in effective learning. Qiu et al.'s (2023) finding about the most effective duration of treatment being between 1-3 months aligns with Dhimolea et al.'s (2022), however Qiu et al. (2023) showed that durations longer than three months as well as those that are less than one month showed smaller effect sizes. Qiu et al. (2023) attributes smaller effect sizes in durations longer than three months to VR fatigue, though more research with statistically significant results need to be shown in order to solidify this argument. In Dhimolea et al. (2023) and Qiu et al. (2023), trends in learning effects for language domains aligned; vocabulary

acquisition showed the most significant and positive effects for learning in comparison to other language domains, with oral communication (or presentation) being second, and reading and writing being either rare or showing negative effects. Dhimolea et al. (2023) and Qiu et al. (2023) expresses the need for more studies measuring VR learning effects in reading and writing. Neither of the reviews nor the meta-analysis included studies with HiVR and L2 pragmatics, which further reiterates the need for a L2 pragmatics VRALL study.

As this dissertation study examines the use of high-immersion VR (HiVR) to teach L2 pragmatics, the findings from Dhimolea et al. (2022) are the most relevant to this study. Dhimolea et al. 's (2022) systematic review includes only empirical studies in which HiVR was used, and it is, at present, the only published VRALL review with HiVR as the object of investigation. In their review, Dhimolea et al. (2022) intended to identify trends in VRALL research and practice, learning outcomes in VRALL research, and learner perceptions of VRALL. Included in this review were 32 journal articles (N=16) and conference proceedings (N=16) published between 2015-2020 (with the majority of publications occurring in 2019 and 2020), many of which were quantitative (N=16) or mixed methods (N=10), had an average sample size of 34 with university aged (18-25 years) participants, and used a variety of HiVR devices (PC-powered, smartphone-based, and standalone), custom-made VR environments (N=16) or pre-existing non language-learning VR content (N= 10) featuring either computer-generated 3D content, 360 degree videos, or a combination of both.

Dhimolea et al. (2022) observed that there are very few VRALL applications for commercial use, and that perhaps this is the reason why most of the VRALL studies in this review used custom-made VR environments. This is a call for more affordable and commercially-available VRALL applications to come onto the market, but likewise, it is a call for more research using the few VRALL applications that are currently available. I addressed this particular call for more research in my dissertation study as I used ImmerseMe, an existing VRALL product.

The languages of the studies reviewed in Dhimolea et al. (2022) include: English, Chinese, Spanish, German, Portuguese, Japanese, French, Arabic, Turkish, and Italian, with English being either the most studied language (N=16) or the most frequent first language of participants (N=8), while Chinese was the second. This finding reiterates a trend that is also found in other VRALL reviews (Qiu et al. 2023; Parmaxi 2020; Huang et al. 2021; Palmeira et al. 2020; Peixoto et al. 2021, Lin & Lan 2015). It is evident that more research is needed for learning outcomes in foreign languages other than English (and with participants whose L1 is not English), however this dissertation study focuses only on learning English as a second language so it does not address this specific gap and it further perpetuates the idea of English being the more popular target language. Just as it does not address the gap in research in the target language of study, this dissertation study also features the same age group (18-25-year-olds) in a university setting and sample size similar to the majority of studies in Dhimolea et al.'s (2022) review, so it does not address the gap in research for HiVR use with higher sample sizes with diverse age groups.

This dissertation does address a specific gap in HiVRALL research, which is that of L2 pragmatics as the language domain of focus. The language domains explored in the studies reviewed in Dhimolea et al. (2022) include vocabulary, oral communication, writing, pronunciation, listening, speaking, and reading, with vocabulary being the most studied (N=16) and reading the least studied (N=1). Interestingly, Dhimolea et al. (2022) identified grammar as a language domain missing from the studies reviewed, however, they did not mention the absence of L2 pragmatics as a language domain. Reasons for not mentioning this could be that L2 pragmatics is a rarely studied language domain (in comparison to other domains), or that the domain of oral communication could be seen as closely related to L2 pragmatics, therefore the authors did not think it was necessary to mention its absence.

Findings from Dhimolea et al.'s (2022) review have also informed the design of this dissertation study. Of the 32 studies reviewed, many of them (N=23) measured learning outcomes and within these studies there was a variety of exposure (e.g., repeated vs. non-repeated) and due to this, Dhimolea et al. (2022) were able to compare learning outcomes of each type of exposure. Dhimolea et al. (2022) identified nine studies with repeated exposure, and fourteen with one-time exposure. All of the repeated exposure studies (N=9) showed positive effects, while only six of the fourteen one-time exposure studies showed positive effects. The remaining eight one-time exposure studies showed negative effects (N=6) and no effect (N=2). This finding is a clear indication that learning outcomes are more likely to be positive with repeated exposure (of two weeks to one semester). Dhimolea et al. (2022) claimed that there is a gap in research for repeated

exposure studies that measure the learning outcomes of communication skills, specifically. For this reason, I decided to conduct a repeated-exposure study of five weeks, with the first week being used for participants to familiarize themselves with the VR technology, and the second to fifth weeks being used for pragmatic tasks.

As for VR and L2 pragmatics, though studies using HiVR to teach L2 pragmatics are presently absent, there have been very few studies done to explore the affordances of (low-immersion) LiVR for learning L2 pragmatics and this area has much potential, as learners can practice developing pragmatic competence without the risk of negative social consequences (González-Lloret, 2019). Taguchi (2022b) states that the only studies that they have identified for VR and interlanguage pragmatics (L2 pragmatics) are those of Sykes (2013), Tang & Taguchi (2020), Tang & Taguchi (2021), and Vilar-Beltrán & Melchor-Couto (2013) and that none of these studies used immersive VR (or HiVR). For this reason, Taguchi (2022b) carried out a research study on the development of pragmatic tasks using immersive VR; this study did not look at how HiVR can be used to teach L2 pragmatics; however, it used a HiVR environment to carry out research in pragmatic development and found that the cognitive demands of VR-based role-plays were different from those computed-based role-plays (participants showed lower fluency in speech and higher use of modifications). Taguchi (2022b) recommends future studies be done in order to see if using immersive VR to teach speech acts produces greater learning outcomes. Thus, this dissertation study would fill the gap for teaching L2 pragmatics with HiVR.

From the literature reviewed here, it is clear that learners tend to have optimistic attitudes towards using VR to learn a language and VR, especially HiVR, which has been shown to provide learners with an environment that helps them feel more confident and relaxed, reducing anxiety. In addition, both LiVR and HiVR have shown great potential in helping learners acquire language and other related skills such as vocabulary, listening, speaking, oral communication, intercultural competence, collaboration, and teamwork. Last, the literature reviewed in this dissertation study affirms that L2 pragmatics can be taught in a classroom setting and that explicit instruction with practice and feedback optimize positive learning effects, and that students in SA contexts (such as the participants in this dissertation study) do not necessarily acquire L2 pragmatics successfully on their own and can benefit from L2 pragmatics instruction. Therefore, this dissertation study addresses a gap in the literature of L2 pragmatics instruction in SA contexts and L2 pragmatics instruction via HiVR.

2.6 Research Questions

Immersive VR as a tool for L2 pragmatic instruction is a timely research topic that calls for attention. The potential findings and pedagogical implications for a study that explores the affordances of immersive VR for L2 pragmatics instruction could inform the practice of a new generation of language instructors and benefit learners who need L2 pragmatics instruction in their contexts. For this reason, this dissertation study seeks to answer the following research questions, as they will serve as a starting point for research in L2 pragmatics learning in an immersive VR environment:

RQ1: Does practicing role-play conversations in an immersive VR environment lead to improvement in pragmatic performance in face-to-face conversations with native speakers?

RQ2: Is practicing role-play conversations in an immersive VR environment more effective for the learning of pragmatics than practicing the role-play conversations between classmates in a traditional classroom setting?

RQ3: What are learner attitudes and perceptions towards completing immersive VR tasks?

CHAPTER 3

METHOD

In this chapter, I explain the scope of the current study and the methodological choices and procedures therein. First, I describe the context of the current study and the participants (section 3.1). Then, I describe the design of the L2 pragmatics instructional treatment with HiVR, the VR software utilized for treatment, and the instruments used to evaluate learning effects and learner attitudes (sections 3.2-3.6). Last, I describe the data collection and analysis of the study and how these were used to answer research questions 1-3 (section 3.7).

3.1.1 Research Context

This study was conducted during the Fall semester of 2023, in intermediate-level communication courses in an Intensive English Program (IEP) for international students at a large, public university in the southwestern region of the United States. Students in

this IEP are conditionally admitted international students who do not have a high enough English proficiency test score (i.e., TOEFL, IELTS, etc.) needed for admission into their undergraduate program of choice, and thus they enroll as full-time students in the IEP with the goal of obtaining full admission. Students in this program receive 21 hours of instruction per week in English for Academic Purposes (EAP), which includes specific instruction in reading and writing, listening and speaking, communication and English for Specific Purposes (ESP) courses.

The IEP consists of six levels (Basic 1 and 2, Intermediate 1 and 2, and Advanced 1 and 2), and students typically spend one, eight-week session in each level. Students who complete all advanced level courses with an average of 80% or higher in each class are granted an English proficiency exam waiver for any undergraduate program at the same university. Upon arrival, new students are given a placement exam by the IEP administration and are placed accordingly. Most students are placed into the intermediate level, making it the most populated level. Students enrolled in the basic and intermediate levels are required to take communication as their elective course, while those in advanced levels take an ESP course of their choice (i.e. English for Engineering, English for Business, etc.). The purpose of the communication course is to help students develop communicative competence in real-life contexts. Being as this study examines the effects of L2 pragmatics instruction on learner pragmatic performance, the communication course appeared to be the ideal course as pragmatic competence forms a significant part of communicative competence (Canale & Swain, 1980; Bachman & Palmer, 1996). L2 pragmatics are not specifically taught in any of the IEP's courses, therefore I saw this

study as an opportunity to integrate L2 pragmatics instruction into the curriculum to add value to the IEP.

Other reasons for having chosen to conduct this study in the IEP's intermediate communication course are that the potential sample size for the intermediate level was larger than that of any other levels, in addition to my inside knowledge and understanding of the curricular needs and challenges of the IEP courses. I have worked as a full-time instructor in this IEP since 2013, and thus am familiar with the needs of the student population and curricular challenges. I knew that this course's curriculum was under constant revision and that program administrators and instructors would welcome experimental instruction to help students improve their oral communication skills.

The intermediate communication courses are held twice per week, for 75 minutes each period. One of the two 75-minute periods is held as a traditional class, while the other 75-minute period is used for a program activity known as "Conversation Club", where students from other class sections of intermediate communication meet in the same room to practice conversation and play games. Students are required to attend this activity for communication course credit. Several local, native-English speaker student employees typically attend these sessions to assist the activities coordinator in conversing with the IEP students. I decided to use the 75-minute Conversation Club period for this study, having the experimental group sections come to the Virtual Reality lab (instead of Conversation Club) to receive VR enhanced L2 pragmatics instruction, and the sections in the control group attend the traditional Conversation Club, where they received L2 pragmatics instruction with traditional, communicative student-student practice activities.

3.1.2 Participants

The recruitment yielded 77 participants, from the intermediate levels in the IEP described above (34 from the two course sections receiving VR treatment, and 43 from three non-VR course sections). Participants are international students who are studying English as a Second Language in the university's Intensive English Program (IEP), and come from diverse linguistic backgrounds. Mandarin and Arabic are the most commonly spoken first languages of the participants in this study, most are within the age range of 18-20 years old, and the gender distribution is nearly equal between males and females (see Table 1).

Table 1

Participant Demographic Information

Participant Characteristics	Number (experimental)	Percent (experimental)	Number (control)	Percent (control)	Total
Male	21	62%	20	47%	41 (53%)
Female	13	38%	23	53%	36 (47%)
18-25 years	25	74%	34	79%	59 (77%)
Over 25 years	9	26%	9	21%	18 (23%)
Arabic	15	44%	14	33%	29 (38%)
Mandarin	14	41%	13	30%	27 (35%)
Japanese	2	6%	12	28%	14 (18%)
Korean	2	6%	0	0%	2 (3%)
Spanish	0	0%	2	5%	2 (3%)
Russian	0	0%	1	2%	1 (1%)
Vietnamese	1	3%	1	2%	1 (1%)

3.2 Description of Instructional Treatment Material

The purpose of this study is to examine the effectiveness of using immersive Virtual Reality (VR) to teach English L2 pragmatics to international students at a US university, so the material selected for VR treatment in the current study was ImmerseMe, an off-the-shelf VR software program designed for language learning and used it in a classroom setting. ImmerseMe features a series of lessons for English language learners of all levels. The themed units feature lessons with communicative role-play tasks, where learners must interact with a pre-recorded video of native English speakers in a 360-degree immersive environment. The software has audio input and output and Artificial Intelligence (AI) is used to interpret and evaluate oral responses from learners. This software can be used on a desktop or with head-mounted devices (HMDs) for a HiVR (fully immersive) experience.

ImmerseMe offers a variety of lessons for basic, intermediate and advanced levels of English (as well as various other languages). ImmerseMe can be used on a desktop or with immersive VR headsets. In the current study, I used ImmerseMe with the HMD “Oculus Rift 2” in order to maximize the affordances of an embodied learning experience in HiVR (Kaplan-Rakowski & Gruber, 2021; Dhimolea et al. 2022).

I selected four lessons of ImmerseMe’s intermediate level units focused on pragmatic tasks. I did use any specific criteria for the selection of lessons other than the fact that they were of the intermediate level and required learners to perform a speech act sequence that they would likely encounter in real life. Each lesson featured a role-play script for learners to read, listen to, and comprehend (see Figure 1). Once learners were

ready to practice the role-play script, they had the option to practice the script orally and in written form by selecting “pronunciation,” “typing,” “spelling,” “translation,” or “immersion.” I required students to complete the “pronunciation” and then the “immersion” practice after achieving a score of at least 80% in the “pronunciation” activity. I chose these two practice activities as they are the most relevant to oral communication.

Before initiating the pronunciation activity, students reviewed the role-play script they would use (see Figure 2). Upon entering the immersive VR environment, students heard the speaker initiate the dialogue and they were prompted to respond by speaking into the microphone. During this practice activity, students were presented with response options in written form on the screen so they could choose the response and then read it aloud. If their oral response was acceptable, then ImmerseMe would congratulate them and the bot would move onto the next part of the scripted conversation. If their response was not acceptable, they would be prompted to try again, with unlimited attempts.

Once learners had successfully completed the pronunciation practice, they were instructed to move onto the “immersion” practice activity. The immersion practice activity is structured similarly to the pronunciation activity, but it is more challenging as learners are required to remember their part of the scripted role-play (the responses do not appear in written form). Once they achieved a score of 80% or higher, they had completed the immersion practice.

As I was facilitating the VR lab, the lab assistants and I monitored student progress by having students raise their hands once they finished the pronunciation

activity and verified that they could move on to the immersion activity. Also, students raised their hands anytime they needed technical assistance or had questions about the activity or procedure. In addition to being present during treatment sessions to facilitate the VR lab and help students navigate the VR environment, I was able to monitor learner activity and progress through the teacher dashboard provided to me by ImmerseMe.

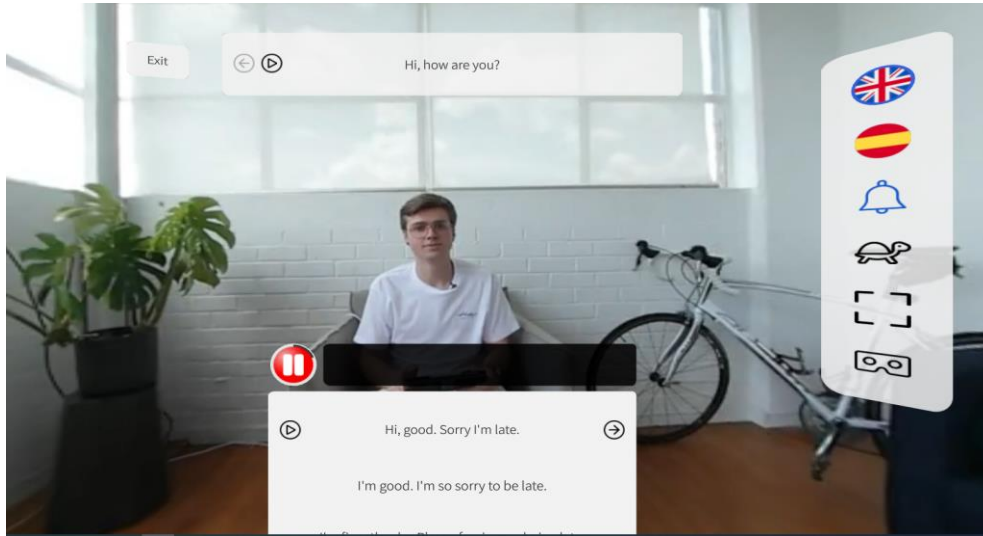
Figure 1

Example of a Role-Play Script in ImmerseMe

The screenshot displays the ImmerseMe interface. On the left, a 'Pronunciation Preview' section contains a 'Lesson Script' with 14 lines of dialogue, each preceded by a green checkmark and a play button icon. The script includes phrases like 'Hi, how are you?', 'Hi, good. Sorry I'm late.', 'It's all good.', 'There was a lot of traffic.', 'I overslept.', 'The metro was delayed.', 'Don't worry, these things happen.', 'I hope you haven't been waiting for too long.', 'Not at all, I've just arrived.', 'Oh that's good, I was getting worried.', 'You're fine. Tracy's also late, she just texted me.', 'Oh, did she say why?', 'No, I guess we'll find out when we see her. Maybe we can grab a drink first.', 'Sounds good. I'll get it, what would you like?', and 'Oh thanks. Just a coke for me, please.' followed by 'Sure, be right back.' On the right side, there is a prominent green 'Start Lesson' button with a play icon and arrow. Below it, a 'Key Vocabulary' section lists words: Traffic, oversleep, metro, delay, text, grab. A 'Key Grammar' section lists verb tenses: Verb tenses > Simple past tense/Verb tenses > Present perfect continuous tense/Verb tenses > Past continuous tense/Verb tenses > Future simple tense/Verb moods > Subjunctive. At the bottom right, there are two 'Download Lesson Script' buttons; the second one specifies '(Including Variations)'.

Figure 2

Example of Pronunciation Activity in ImmerseMe



3.3 Design of Study

Before the start of the Fall 2023 semester, I met with the program director, curriculum coordinator, activities coordinator, and course teachers of the IEP to discuss the design of the study and the curriculum modifications that would need to take place so that all students in the intermediate communication courses could benefit from L2 pragmatics instruction. I was given permission to modify the curriculum of all intermediate level communication course sections (there were five sections in total).

The major modification that was done to the intermediate communication course curriculum was that during the designated Conversation Club activity time (one 75-minute period per week), the two sections selected for VR enhanced pragmatics instruction were required to go to the VR lab instead of the Conversation Club activity room, and the remaining three sections of non-VR pragmatics instruction were required

to attend class in the conversation club activity room where they would receive the same lesson in pragmatics but without VR. The reason for the control group being made up of three sections instead of two or two and one half is that there was limited space and equipment in the VR lab; only 20 headsets (or HMDs) were available at a time, so I had to select the two least populated sections of intermediate communication courses to form the experimental group.

The sections in the control group had between 20-30 students per section, while the sections chosen for VR treatment had between 15 and 18 students. All students received treatment regardless of participation in the study. I planned the lessons and created instructional materials for all treatment sessions and facilitated VR lab, and the activities coordinator facilitated the control group sessions in the designated classroom and class time for the Conversation Club activity. This design ensured that students would receive the same instruction with the same materials, making the VR-enhanced practice the only difference. I met with the activity coordinator on a weekly basis to ensure that she understood the materials and could follow the lesson plan accordingly.

As this is a quasi-experimental, mixed-methods study, I divided participants into two separate groups (experimental and control) and administered a series of pre and posttests to both groups and an attitudinal questionnaire to the experimental group. While both groups received the treatment of L2 pragmatics instruction, the experimental group experienced VR-enhanced L2 pragmatics lessons, while the control group experienced L2 pragmatics instruction in a traditional classroom setting without VR. The length of the data collection period was six weeks, and the treatment took place during weeks 2-5 of

the study (see Table 2). The first week was an introductory and pretest week and each of the following four weeks were used to learn a different speech act sequence. The speech act sequences that were taught during this treatment were: (1) Declining an invitation, (2) Apologizing for being late, (3) Asking for a favor, and (4) Asking about private life politely. The order in which these lessons were presented was random and not based on any specific criteria.

Students in both groups received the same instructional treatment (explained below) of 75 minutes each treatment period, practicing the speech act sequences as role-plays. Students in the control group practiced their role-plays with their classmates, while students in the experimental sections practiced the same role-plays in an immersive VR environment. Therefore, practicing the role-plays in a VR environment on ImmerseMe was the only difference in instructional treatment between the two groups, hence the modality (face-to-face vs. VR) and individual feedback learners received during their practice activities were the independent variables.

During each treatment session, four lab (experimental group) or classroom assistants (control group) were present in each group. The assistants were domestic student employees and native speakers of English who were either working in the VR lab to assist with technical difficulties (experimental group) or as Conversation Club activity assistants in the IEP (control group). This circumstance positioned the classroom assistants as ideal role-play test partners for participants in this study, as they attend the same university as the classroom assistants, are in the same age group, and represent a sample of the native speaker domestic student population that participants would interact

with in the context of this study. For this reason, the domestic student classroom assistants partnered with participants during the last 10-15 minutes of each session and audio-recorded their open role-play tests.

Table 2

Timeline

Week Number	Research Activity
Week 1	Recruitment of Participants
Week 2	Introduction and Pretests 1-4
Week 3	Task 1 Treatment + Task 1 Posttest (Declining an Invitation)
Week 4	Task 2 Treatment + Task 2 Posttest (Apologizing for Being Late)
Week 5	Task 3 Treatment + Task 3 Posttest (Asking for a Favor)
Week 6	Task 4 Treatment + Task 4 Posttest (Asking about Private Life Politely)
Week 7	Attitudinal Questionnaire

Note. The weeks listed are the first seven weeks of an eight-week session.

Experimental and control groups received instruction on the same pragmatic task each week during the VR lab (and non-VR Conversation Club). I facilitated the VR lab while the activities coordinator facilitated the non-VR class session/Conversation Club. I designed the lesson plans for both groups and met with the activity coordinator on a weekly basis to review the lesson plan and materials for that week. The lessons/treatment sessions were each 75 minutes in length, with the first 15-20 minutes of class being dedicated to explicit metapragmatic instruction (discussions to raise metapragmatic

awareness, explanations of expected conversational moves and example dialogues), followed by a productive practice task known as *closed role-plays* (scripted scenarios with assigned dialogues and roles) for 45-50 minutes, with the last 10-15 minutes of class being used for posttests (prompt-based, face-to-face, open role-plays).

3.4. Rationale for Instructional Treatment

I designed the treatment lessons under several assumptions: (1) L2 pragmatics can be learned are teachable in a classroom setting (Kasper & Rose, 1999), and that (2) explicit, form-focused instruction (Ellis, 2005) renders higher gains in learning than implicit instruction, and that (3) explicit L2 pragmatics instruction renders superior learning outcomes to implicit instruction (Taguchi, 2015; Jeon & Keya, 2006; Takahashi, 2010a; Plonsky & Zhuang, 2019). I employed the explicit method of L2 pragmatics instruction by giving a metapragmatic explanation about the speech act sequence of focus, and then requiring students to practice said sequence after receiving explanation, as recommended by Kasper (2001) and employed in studies that showed larger learning effects for explicit instruction that includes explicit teaching, practice, and feedback, over implicit instruction (Plonsky & Zhuang, 2019; Taguchi, 2015).

I started each lesson with a warm-up discussion to raise metapragmatic awareness among students (as done in Alcón-Soler, 2007) and followed with a metapragmatic explanation of each speech act sequence by highlighting the expected conversational moves of the sequence, reasons for these moves, and I provided handouts with an example role-play of the speech act sequence of focus. The closed role-play practices were considered to be the production aspect of the lesson, as seen in Félix-Brasdefer

(2008) and House (1996). Both experimental and control groups used the same role-play scripts from ImmerseMe to learn and practice the speech act sequences of focus. Students in the VR treatment group received instant, individual feedback from the ImmerseMe software as they practiced in the VR environment, while students in the control group got limited to no feedback from their peers as they practiced the role-plays in pairs with other students.

The length of instructional treatment for each speech act sequence was 75 minutes, and 45-50 minutes of the 75 minutes was used for production-skill practice. I chose this length due to institutional and material constraints of my study. I wanted students to learn a new speech act sequence, (instead of practicing the same one), each treatment session in order to avoid redundancy and boredom which can lead to VR fatigue as mentioned in Qiu et al., (2022). Also, the ImmerseMe role-play scripts were fairly short and there was only one script per task/lesson. If I had required students to repeat the same role-play task practice in the VR environment each week, the risk of VR fatigue would have been higher. In previous studies reviewed by Takahashi (2010a) and Taguchi (2015), treatment lengths varied from one class session to multiple hours of practice per week over an extended period of time and these treatment lengths were decided based on the context of the study, not empirical evidence of treatment length or quantity of practice. Taguchi (2015) explains in her review that though some research studies show that longer length of treatment leads to higher gains in learning, more research is needed to prove empirical significance. In addition, Taguchi (2015) states that “simple length of time may not help indicate the quality and intensity of practice” and

that “Because there is no information about how much time is actually spent in processing the target pragmatic features, a plain comparison of instructional times may not be so fruitful” (p.32). Therefore, in terms of length of instruction and practice, there is a lack of strong, empirical evidence proving the effectiveness of a certain quantity of time for the learning of a speech act sequence, hence, the length of time for instruction and practice in this dissertation study was based on practicality and context only.

3.5.1 Research instruments

The primary instrument of measurement used in this dissertation study was open role-plays (as defined in Kasper & Dahl, 1991), where students read a prompt and responded orally by interacting with an interlocutor as they would in real life. Role-plays have been used in various L2 pragmatics studies in quasi-experimental, educational research contexts to assess learner pragmatic knowledge or performance, and they are commonly used in conjunction with other types of tests such as Discourse Completion Tests (DCTs) and verbal reports (Nguyen, 2019; Félix-Brasdefer, 2010). Support for using role-plays in studies such as this dissertation study is described in Taguchi’s (2015) review, which states that role-plays are different from DCTs in that they require learners to use multiple levels of processing in order to perform a task, as they are required to participate in an impromptu, interactional dialogue in which they co-construct a discourse. This creates a scenario which simulates real-life impromptu interactions in the wild. In this dissertation study, participants’ role-play partners were domestic students whose L1 was English, making these role-plays even more similar to real-life, face-to-face interactions students would have on campus. Taguchi (2015) reiterates that role-

plays “could be a more accurate representation of learners’ performance, determining whether learners can use their pragmalinguistic knowledge to accomplish higher-level objectives such as negotiation, mutual understanding, and interactional goals” (p. 13).

According to Félix-Brasdefer (2010), pragmatic competence is a hypothetical construct that researchers cannot measure directly due to several practical challenges, therefore researchers are bound to indirect assessment measures (such as DCTs, role-plays and verbal reports). Practical challenges to assessment include control of learner variables (age, gender, socioeconomic background, etc.) and the ability to collect data for frequent observations of interactions with native speakers outside of the classroom (Félix-Brasdefer, 2010). I addressed some of the typical concerns with validity and reliability described in Félix-Brasdefer (2010).

Regarding content validity, I wrote the role-play prompts to include a description of the scenario, context and participants’ relationship to their interlocutor, and used gender neutral language so that participants could adapt the prompt to their gender identity. I provided a rich description of the context and scenario in the prompts and prompt instructions in order to increase the content validity of the instrument, following Félix-Brasdefer and Bardovi-Harlig’s (2010) example. Félix-Brasdefer (2010) affirms that role-plays have been observed as moderately reliable and valid instruments in various L2 pragmatics studies over time, but that they should only serve the purpose of testing for research and not be used as tools to assess L2 proficiency.

I addressed item validity by ensuring that the prompts were directly related to the target task of measurement, and that the rubrics used to evaluate pragmatic performance

were descriptive and specific to measuring two primary aspects of pragmatic performance, which are appropriateness and grammar (Taguchi, 2012). I adapted these rubrics from Taguchi's (2012) study to increase construct validity. In this study, I chose to include Taguchi's (2012) grammar rubric as a separate measure for several reasons (other than the fact the Taguchi (2012) included it in her study). Bardovi-Harlig (2003) states that "pragmatic development cannot proceed completely independently of grammatical development" (p. 25). An expression must first be acquired and one must know syntactical placement of the expression before one learns when, where, how and with whom to use said expression, therefore, grammar must be considered when looking holistically at learner interlanguage development. Bardovi-Harlig (2003) reviewed two longitudinal studies in which found that learners with greater grammatical development were able to communicate using fewer repairs and other mitigation strategies; though these are not conclusive, they do serve as an indication that when learners acquire the pragmalinguistics needed, they can deploy more cognitive resources and pay attention to the sociopragmatic cues and dynamic changes during interaction in the real world.

Reliability was addressed through the rating of role-play tests. For rating, I utilized the rating scale approach described by Taguchi and Li (2020) and Taguchi (2012). One benefit of using a rating scale is that it has potential for a high level of construct coverage because they can include a more dynamic and comprehensive evaluation of abilities (Taguchi & Li, 2020), and this approach is flexible in that researchers can modify the scales to fit the needs of their studies and contexts. I followed Li et al. (2019) and Taguchi and Li's (2020) recommendation that test raters undergo

rater training. The other rater and I had an anchoring session prior to rating the tests and reconciled scores post rating. Interrater reliability is addressed in the results section in this study.

Given the methodological considerations mentioned above, I wrote the prompts with a description of the context and task that international students and domestic students were required to carry out. I designed the pre and posttest tasks to be equally or slightly more challenging than the scripted one's participants practiced during the session. I also considered the reality of the students when writing the prompts - the prompts reflect typical scenarios in which university students in the US would communicate with each other. For example, the "Asking for a Favor" lesson, the scenario described in the posttest prompt states that the international student and domestic student are roommates, and the international student needs help lowering their luggage from their apartment to the car, and then they need a ride to from their apartment to the airport; this prompt, for example, positioned the international student to request a face-threatening favor. A scenario such as this could realistically occur in the participants' context studying abroad at a US university (see Appendix B for all prompts).

This instrument was critical in assessing students' ability to perform pragmatic tasks in a conversational exchange with a domestic native speaker their age, in a US university social context. As these role-plays were open and prompted (but not scripted), it was assumed that the native speaker student employees would respond to their international student partners in the same way they would respond in real life. I acknowledge that the prompts helped create a conversation that would prompt students to

perform the task they learned that day and that real-life occurrences of these conversations may have gone differently, making these role-play tasks inauthentic. I also acknowledge that the domestic native speaker students who served as role-play partners were not a homogenous group and that the variability in gender identity, ethnic origin, willingness to engage with L2 users of English or international students, intercultural competence, among other individual characteristics, were not accounted for. However, given the nature of the study and context in which it was conducted, the open role-play tests were designed to simulate real life, naturally occurring conversations between international and domestic students on a US university campus, and this made role-plays a more authentic choice of instrument in comparison to DCTs.

3.5.2 Attitudinal questionnaire

The second instrument used in this study was an attitudinal questionnaire. The purpose of the attitudinal questionnaire was to obtain data regarding student experiences using VR, hence it was only administered to the sections that received VR treatment. Another purpose of this questionnaire was to gather information for the institution and IEP administration, to see if students in the program would be interested in using VR in future semesters, and challenges for implementation. The questionnaire consisted of nine questions: four Likert-scale, two multiple-choice and three open-ended questions (see Appendix C).

3.6 Data Collection Timeline

Data collection took place over a period of six weeks, within an eight-week session. During the first week of the Fall A session, I received IRB approval from my

institution (see Appendix B) and then recruited participants for the study by emailing them through their course LMS, and visiting them in person during class time. I completed the recruitment and consenting process by the end of the first week of the Fall A session. The second week of the session was the first day in the VR lab (and Conversation Club activity) and I used this time to complete the role-play pretests and introduce students to VR so that they could learn how to use it prior to starting treatment.

Instructional treatment began the third week of the session and ended the fifth week, which resulted in four treatment periods. The role-play posttests were given during the last 10-15 minutes of the weekly lab immediately after instruction and practice during the four weeks of treatment. I assigned the attitudinal questionnaire to the course sections in the experimental group during week six of the session on the intermediate communication course LMS. The seventh week of the session consisted of a celebratory activity, and the eighth week was reserved for final exams. Hence, I fit four weeks of treatment into the eight-week session of Fall A, 2023 and collected the data (see Table 2).

3.7 Data Collection Procedures

The methodological approach to data collection in this study is mixed methods, but mostly quantitative. To answer research questions one and two, I administered pre and posttest role-plays and rated them for appropriateness and grammar. To answer research question three, I administered an attitudinal survey to the experimental group and analyzed the results in the context of my study. Throughout the duration of this study, I collected pre and posttest role-play recordings on a weekly basis as they were

given, and I collected responses to the attitudinal questionnaire during the last week of the study.

The recorded role-plays were stored on eight separate digital audio recorders (four for each group). The digital audio recorders were my personal property and I purchased them only for the purpose of this study. For each lab session, I brought the digital recorders with me, delivered them to the classroom/lab assistants in each group (experimental and control), and collected them immediately after each treatment session. This was easy to do as the treatment sessions happened concurrently and in the same building (in different rooms). I stored the digital recorders in a drawer in my office and secured them with a lock and key.

At the end of each week of treatment, I uploaded the recorded role-plays to my personal, institutional computer, and labeled each recording according to the specific task/test code and identification number assigned to each consenting participant. Then, I edited the first few seconds of each audio recording where students said their names. Then, I uploaded these files into folders in my secured personal, institutional Google drive account. The recordings of students who did not consent to participate in the study were deleted as they were identified. Each week, I deleted recordings from the devices after uploading them to my personal computer and saving them to my Google drive. After labeling each recording with a code for the test and student ID number, I edited the first few seconds of each audio recording where participants state their name so that they would not be identifiable to the test raters. Then, I organized the recordings into a folder and created a Google sheet for each rater. The Google sheets included the code of each

test (student ID number and test number) with a hyperlink to their corresponding audio recordings, a column for appropriateness scores, a column for the grammar scores, with appropriateness and grammar rating rubrics pasted onto the sheets for easy reference.

3.8 Data Analysis

The role-play recordings were rated by two raters: myself and a colleague who work as instructors in the same IEP. My colleague is a native speaker of American English and I am a native English/Spanish bilingual; we share similar educational backgrounds and work experience as TESOL professionals. In addition, we both have extensive experience in rating standardized foreign language tests inside and outside of our program. To rate the role-play tests, we used two separate rubrics with a rating scale of 0-5. One rubric was used to measure appropriateness and the other was used to measure grammar; both rubrics featured scales of 0-5, with zero being no response and five being nearly perfect or native-like (see Appendix C).

Before rating the tests (1-3 days prior), my colleague and I met in-person for a one-hour anchoring session, in which we agreed on how to use the rubrics and what we would expect for each rating on the scale. As I had coded the hyperlinked test labels with letters and student ID numbers on the rating Google sheets, the second rater was unable to decipher between which tests were pre and which were post. Likewise, when I was rating the tests I could not immediately identify which tests were pre or post without referencing my coding sheet.

I rated all tests first in order to avoid rating bias from viewing the second rater's scores. The second rater did not have access to my Google sheet with pre and posttest

scores in order to avoid rating bias as well; I was the only one with access to both sheets, and after the second rater finished entering scores into her Google sheet, I created a new Google sheet to display both raters' scores for comparison (which only I had access to). I then compared scores and identified which scores were two or more steps off (for example, 2 vs. 4, or 1 vs. 5, etc.), and met with the second rater to reconcile these scores. We listened to each audio recording twice, re-scored, and discussed our scores until we reached an agreement on one, final, numerical score.

For the scores that were only one step off, I calculated the final score by taking the average of the two scores (for example, if rater 1 gave a 4 and rater 2 gave a 5, then the final score would be 4.5). Finally, I created a new Google sheet with all pre and posttest data for statistical analysis. After obtaining all final pre and posttest scores, I did not calculate the Cohen's Kappa value for interrater reliability for these scores because the averaging of scores that were one step off would have created artificially low Kappa values. Instead, frequency of agreement is displayed in Chapter 4 (see Table 2).

Next, to answer Research Question 1 (Does practicing role-play conversations in an immersive VR environment lead to improvement in pragmatic performance in face-to-face conversations with native speakers?), the pre and posttest scores and role-play scores for appropriateness and grammar from the experimental group were analyzed using the Wilcoxon signed-rank test in SPSS software. I used this non-parametric test for the data analysis pertaining to RQ1 in order to measure differences in paired samples that are not necessarily normally distributed.

For Research Question 2 (Is practicing role-play conversations in an immersive VR environment more effective than practicing the role-play conversations between classmates in a traditional classroom setting?) I analyzed the gain scores (pre and posttest difference) for each set of tests by conducting the Mann-Whitney U test in SPSS software. I applied this non-parametric test to see if there were differences in pre-posttest gain scores of two independent groups (experimental and control) that have data that are not normally distributed. In addition, I calculated the descriptive statistics, including the mean rank scores, the p-value and the effect size in order to identify any trends.

Last, I administered an attitudinal questionnaire to answer Research Question three (What are learner attitudes and perceptions towards completing immersive VR tasks?) on the students' course LMS. The questionnaire consisted of nine questions (four Likert scale, two multiple choice, and three short answer questions). For questions 1-4, I calculated the means and standard deviations of the four-point Likert scale responses. Then, I assessed the reliability of these questions by computing Cronbach's Alpha. For questions 5 and 6, I calculated the frequency of responses to identify majority opinions and perspectives. For questions 7-9, the short answer questions, I read participant responses and coded them for common themes. Then, I shared my coding and participant responses with a colleague (the same colleague who rated the role-play tests) and she verified my coding by reading the responses and counting the number of responses I coded per theme. Discrepancies in theme coding and counting were resolved by meeting in person to discuss them until we reached complete agreement.

3.9 Chapter Summary

In this chapter, I described the design of the study, instructional treatment, and instruments used. The methodological rationale for instructional treatment and data collection instruments was discussed, and the methods for scoring the role-play pre and posttests and coding the attitudinal questionnaire were explained. The results from the data collection are presented in the following chapter (Chapter 4).

CHAPTER 4

RESULTS

This chapter displays the results for the statistical analysis for the data collected in this study. The quantitative analysis was done to answer research questions 1 and 2 (section 4.1), while a qualitative analysis was done to answer research question 3 (section 4.2). As the data for pre and post role-play tests were largely non-normal, I decided to use non-parametric measures. Therefore, the Wilcoxon sign-ranks test and the Mann Whitney U tests were selected for the data in this study because they are non-parametric tests that are appropriate for non-normal data which have modest sample sizes. Results from these tests are displayed in tables below. Interrater reliability for the role-play pre and posttests was considered and is displayed in percentage of agreement. The qualitative survey results are reported using descriptive statistics, frequencies and coding for themes. Last, a summary of results is provided (section 4.3).

4.1 Quantitative Analysis

The role-play pre and posttests in this study (see Appendix D for test scores) were rated by the same two raters, and each recorded role-play was evaluated by each rater using two rubrics with a 0-5 ranking scale, one for appropriateness and one for grammar (see Appendix E), as each of these are aspects of pragmatic competence. For rating agreement, if the scores were within one step of another we averaged them, if there was more than one, we consulted with one another until we agreed on one score. Tables 1.1 and 1.2 below show the number of tests rated, the number and percentage of scores that were averaged due to being one step off, the number and percentage of reconciled scores,

and the number of scores with perfect agreement prior to averaging or reconciliation. The raters demonstrated agreement for the majority of scores in appropriateness and grammar rating of all tests, with 67% being the lowest percentage of perfect agreement and 89% being the highest (see Tables 3 and 4).

Table 3

Interrater Agreement for Appropriateness Role-Play Tests

Appropriateness Tests	Number of Tests	Averaged Scores	Reconciled Scores	Perfect Agreement
Pretest Task 1	70	9 (13%)	1 (1.4%)	60 (86%)
Posttest Task 1	60	17 (28%)	0	43 (72%)
Pretest Task 2	70	9 (13%)	1 (1.4%)	60 (86%)
Posttest Task 2	59	16 (27%)	0	43 (73%)
Pretest Task 3	70	10 (14%)	3 (4.5%)	57 (81%)
Posttest Task 3	61	6 (10%)	0	54 (89%)
Pretest Task 4	70	14 (20%)	0	56 (80%)
Posttest Task 4	54	12 (22%)	0	42 (78%)

Note. Perfect agreement refers to the number and percentage of total scores with perfect agreement prior to averaging scores that were one step above or below and before reconciliation of scores that were two or more steps off.

Table 4*Interrater Agreement for Grammar Role-Play Tests*

Grammar Tests	Number of Tests	Averaged Scores	Reconciled Scores	Perfect Agreement
Pretest Task 1	70	8 (11%)	0	62 (89%)
Posttest Task 1	60	20 (33%)	0	40 (67%)
Pretest Task 2	70	10 (14%)	0	60 (86%)
Posttest Task 2	59	18 (31%)	0	41 (69%)
Pretest Task 3	70	8 (11%)	0	62 (89%)
Posttest Task 3	61	12 (20%)	0	49 (80%)
Pretest Task 4	70	9 (13%)	0	61 (87%)
Posttest Task 4	54	11 (80%)	0	43 (80%)

Note. Perfect agreement refers to the number and percentage of total scores with perfect agreement prior to averaging scores that were one step above or below and before reconciliation of scores that were two or more steps off.

4.2 Research Question 1

The first research question asks: Does practicing role-play conversations in an immersive VR environment lead to improvement in pragmatic performance in face-to-face conversations with native speakers? To answer this, I obtained the descriptive statistics for pre and posttests of appropriateness and grammar in each group (see Tables 2.1 and 2.2), and then applied a series of Wilcoxon signed-rank tests to compare pre- and posttest mean scores in the role-play tests for appropriateness and grammar. These non-parametric tests were used due to the ordinal nature of the scores (0-5) and that the pre – posttest measures were on paired groups. Results for appropriateness will be reviewed followed by those for grammar (see Tables 5 and 6).

In the descriptive statistics shown in Table 5, it can be observed that for the appropriateness measure, participants in the experimental group had higher average posttest scores than the control group in Tasks 1-3 and slightly lower pretest scores in all tasks.

Table 5
Pre and Posttest Descriptive Statistics for Appropriateness

Test	Group	Pretest					Posttest				
		N	M	SD	Min	Max	N	M	SD	Min	Max
Task 1	Exp.	32	3.17	0.64	2	4	27	3.72	0.80	2	5
	Ctrl.	35	3.47	0.72	2	5	32	3.53	0.99	2	5
Task 2	Exp.	32	3.18	0.76	1	1.45	27	3.55	0.64	2	4.5
	Ctrl.	35	3.34	0.63	2	5	32	3.50	0.67	2	5
Task 3	Exp.	32	3.06	0.85	1	5	30	3.65	0.90	2	5
	Ctrl.	35	3.28	0.65	2	4	30	3.86	0.73	2	5
Task 4	Exp.	32	2.90	0.74	2	4	25	3.22	0.66	2	4
	Ctrl.	35	3.11	0.83	1	5	28	3.07	0.81	1	4

Note. Abbreviations for the experimental and control group are exp. and ctrl.

In descriptive statistics in Table 6 it can be observed that for the grammar measure, participants in the experimental group had slightly higher average posttest scores than the control group in Tasks 1-3 and slightly lower pretest scores in all tasks.

Table 6

Pre and Posttest Descriptive Statistics for Grammar

Test	Group	N	Pretest				Posttest				
			M	SD	Min	Max	N	M	SD	Min	Max
Task 1	Exp.	31	2.93	0.30	2.5	4	27	3.14	0.58	2	4
	Ctrl.	36	3.08	0.43	2	4	32	3.09	0.57	2	5
Task 2	Exp.	31	2.75	0.48	2	4	27	3.18	0.44	2	4
	Ctrl.	36	2.95	0.53	2	5	32	3.12	0.52	2	4
Task 3	Exp.	31	2.87	0.36	2	4	29	3.20	0.50	2	4
	Ctrl.	36	3.11	0.34	3	4.5	30	3.05	0.47	2	4
Task 4	Exp.	31	2.75	0.33	2	3	24	3.02	0.31	2	3.5
	Ctrl.	36	3.04	0.70	1	5	30	3.05	0.44	2	4

Note. Abbreviations for the experimental and control group are exp. and ctrl.

Overall, Table 7 shows that the experimental group had significant gains in appropriateness scores from pre- to post-test in all but one of the tasks. In contrast, the control group only had one significant increase from pre- to post-test. The experimental group showed significant differences for Tasks 1-3, with Task 3 being the most significant ($p=0.014$). Interestingly, Task 3 was the only task in which the control group showed a significant difference ($p=0.005$), performing equally to or slightly higher than the experimental group. The magnitude of the effect sizes for gains made by the experimental group are moderate to large for all tasks, while the effect sizes of the control group are small. Holistically, in terms of the appropriateness measure, it can be observed that the experimental group outperformed the control group, which supports the hypothesis that learners improve their pragmatic performance in face-to-face conversations with native speakers due to practicing speech act sequences (as role-play conversations) in an immersive VR environment.

Table 7

Wilcoxon Signed-Ranks Tests for Appropriateness in Experimental and Control Groups

Test	Group	N	Mean Rank	z	W	p	Effect Size
Task 1	Experimental	27	10.07	2.26	151	0.02	0.43
	Control	32	15.03	0.53	225.50	0.59	0.09
Task 2	Experimental	27	10.77	2.16	161.50	0.03	0.41
	Control	32	12.14	1.61	133.50	0.10	0.28
Task 3	Experimental	30	12.82	2.46	218	0.01	0.45
	Control	30	10.62	2.84	180.50	0.00	0.01
	Experimental	25	12.13	1.72	194	0.08	0.34

Task 4	Control	28	7.19	-0.92	57	0.35	0.17
--------	---------	----	------	-------	----	------	------

Note. P-values are two-tailed. Effect size was obtained from the r statistic.

Overall, Table 8 shows that the experimental group showed significant increases in grammar in all but one task. The control group saw no significant increases. The results of the Wilcoxon sign-ranks test show moderate to large effect sizes for all tasks for the experimental group for grammar. Likewise, with the exception of task 1, all p-values for the experimental groups' tasks show a significant pre and posttest difference, while none of the p-values of the control group reach significance. Tasks 2 ($p=0.001$) and 3 ($p=0.009$) had the most significant pre-posttest difference within the experimental group, and the magnitude of these differences is observed in moderate to large effect sizes. Overall, the experimental group outperformed the control group in all tasks, including the only task where no significant difference was observed for the experimental group (Task 1). These findings suggest that the grammatical aspect of students' pragmatic performance in face-to-face conversations with native speakers improved after practicing speech act sequences (as role-play conversations) in an immersive VR environment.

Table 8*Wilcoxon Signed-Ranks Tests for Grammar in Experimental and Control Groups*

Test	Group	N	Mean Rank	z	W	p	Effect Size
Task 1	Experimental	27	8.21	-1.32	115	0.18	0.25
	Control	32	10.64	0.34	74.50	0.73	0.06
Task 2	Experimental	27	10.18	3.42	193.50	0.00	0.65
	Control	32	7.25	1.56	87	0.11	0.27
Task 3	Experimental	29	10.47	2.60	157	0.00	0.42
	Control	30	8.06	-0.60	64.50	0.54	0.14
Task 4	Experimental	24	8.14	2.54	114	0.01	0.51
	Control	30	8.28	0.34	74.50	0.73	0.06

Note. P-values are two-tailed. Effect size was obtained from the r statistic.

4.3 Research Question 2

The second research question asks: Is practicing role-play conversations in an immersive VR environment more effective than practicing the role-play conversations between classmates in a traditional classroom setting? I applied a series of Mann-Whitney U tests to compare gain scores across role-play tests for appropriateness and grammar. These non-parametric tests were used due to the ordinal nature of the scores (0-5), there being two independent groups compared, and that the data was not normally distributed and equal variances cannot be assumed. Results for appropriateness will be reviewed followed by those for grammar (see Tables 9 and 10).

In reference to the appropriateness measure, the Mann-Whitney U test (see Table 9) indicated highly significant gains for the experimental group over the control group in Tasks 1 and 4. The task in which the highest gains were observed is Task 4 ($p=0.05$). The magnitude of the gains is observed in large effect sizes for Tasks 1, 2 and 4. This finding serves as strong evidence to support the hypothesis that, in terms of appropriateness, learners who receive individualized practice and feedback for speech act sequences in an immersive VR environment show higher improvement in pragmatic performance in comparison to learners who practice speech act sequences with their peers in a traditional classroom environment.

Table 9

Mann-Whitney U Test Results for Role-Play Test Gain Scores for Appropriateness

Test	Group	N	Mean Rank	U	Z	p	Effect Size
Task 1	Experimental	27	34.98	297.50	-2.08	0.03	0.69
	Control	32	25.80				
Task 2	Experimental	27	31.43	366.50	-0.83	0.40	0.88
	Control	31	27.82				
Task 3	Experimental	30	29.22	488.50	0.57	0.56	1.1
	Control	30	31.78				
Task 4	Experimental	24	30.94	229.50	-1.99	0.05	0.68
	Control	28	22.70				

Note. P-values are two-tailed. Effect size was obtained from the r statistic.

In reference to the grammar measure, the Mann-Whitney U test (see Table 10) indicated significant gains for the experimental group over the control group in Tasks 1 and 3. The highest significance shown is that of Task 1 ($p=0.05$). The magnitude of the observed gains is evident in the reported effect sizes; effects for all tasks are large. This is evidence that, when compared to learners who practice speech act sequences (as role-plays) in a traditional classroom environment, students who practice the same role-plays in an immersive VR environment (where individual practice and feedback occur), show significant improvement in the grammatical aspect of their pragmatic performance.

Table 10

Mann-Whitney U Test Results for Role-Play Test Gain Scores for Grammar

Test	Group	N	Mean Rank	U	Z	p	Effect Size
Task 1	Experimental	17	27.59	160	-1.921	0.05	0.67
	Control	28	20.21				
Task 2	Experimental	19	28.21	186	-1.87	0.06	0.69
	Control	28	21.14				
Task 3	Experimental	17	26.21	132	-2.16	0.03	0.62
	Control	25	18.30				
Task 4	Experimental	18	25.33	165	-1.55	0.12	0.73
	Control	25	19.60				

Note. P-values are two-tailed. Effect size was obtained from the r statistic.

4.2 Qualitative Analysis

The final research question asks: What are learner attitudes and perceptions towards completing immersive VR tasks? To answer this, I examined the experimental group participants' attitudes and perceptions on their experience learning speech act

sequences with VR by issuing an attitudinal survey to the experimental group. The survey consisted of nine questions: a Likert-scale ranging from 1-4, two multiple choice, and three short answer questions. Results are displayed in Tables 11 and 12, and Figures 3 and 4.

Table 11 reports average ratings and standard deviations for each of the Likert scale questions (questions 1-4), with all 25 scoring above 3 on a 1-4 scale. The Likert scale questions were assessed for reliability by computing Cronbach's Alpha and obtained a value of 0.68, indicating moderately high internal consistency among the questionnaire items.

Table 11

Survey Items with Likert Scale Ratings (Questions 1-4)

Survey Item	M	SD
1. My experience using Virtual Reality in the English classroom was great!	3.16	2.74
2. The Virtual Reality activities helped me practice speaking English.	3.32	2.83
3. I feel more confident in my English-speaking skills after completing the Virtual Reality activities.	3.04	2.59
4. Global Launch should offer Virtual Reality to all students, every session.	3.04	2.65

Tables 12 and 13 display the frequency and percent of responses to multiple-choice questions 5 and 6. For Question 5, nearly half (48%) of respondents selected “Asking for a Favor” as the lesson they liked the most, making it the most popular lesson among the four. Question 6 shows a nearly equal distribution among responses showing a slight majority preference for an avatar in an animated environment (40%).

Table 12*Distribution of Multiple-Choice Responses (Question 5)*

Which of the following VR lessons did you like the most?		
Responses	Frequency	Percent
Apologizing for being late	5	20.0
Asking about private life politely	4	16.0
Asking for a favor	12	48.0
Declining an invitation	4	16.0
Total	25	100.0

Table 13*Distribution of Multiple-Choice Responses (Question 6)*

Did you like the real-life videos in the software we used, or would you prefer a game-like experience with avatars?		
Responses	Frequency	Percent
I do not have a preference.	6	24.0
I like the real-life software we used in class.	9	36.0
I would prefer an avatar(cartoon) with an animated (game) environment.	10	40.0
Total	25	100.0

Figure 3 displays the coded themes and number of responses for each theme. The two most frequent themes were Novelty (6 responses) and Feedback and Practice (5

responses). Novelty refers to comments that respondents made about VR being new, different, memorable, and exciting. Feedback and Practice refers to comments that participants made about the VR lessons being a good source of practice and feedback for the task at hand.

Figure 3

Frequency of Coded Themes for Short-Answer Question 7

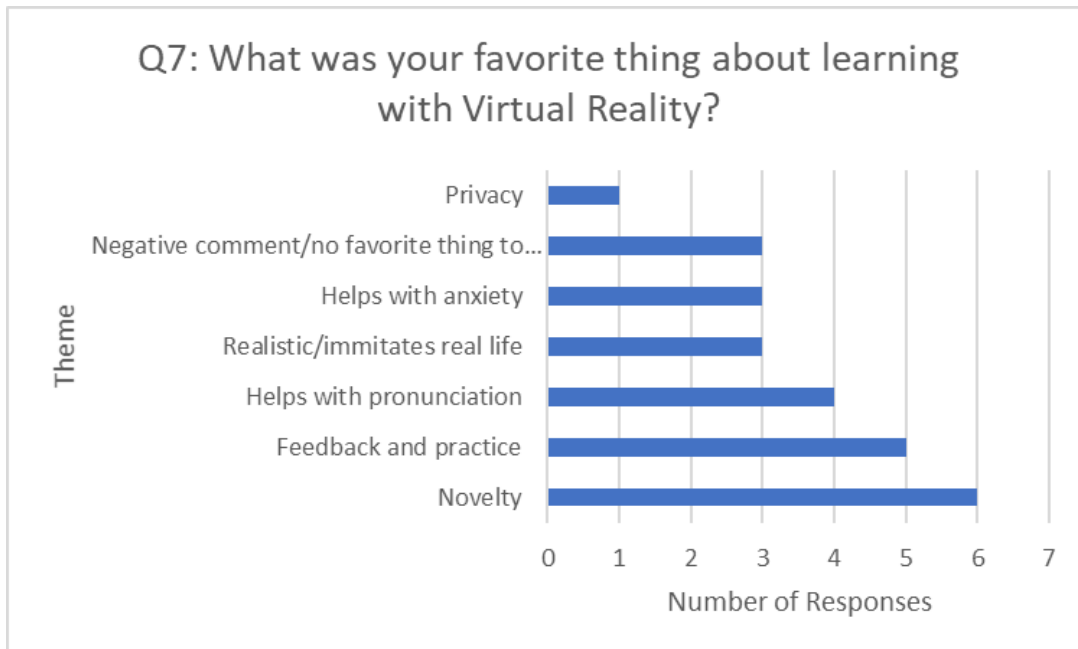


Figure 3 displays the coded themes and number of responses for each theme in Question 7. The most frequently selected theme (12 respondents) was “Issues with Audio Input/Output.” This theme emerged from comments respondents made about not being able to hear the person who was speaking to them in the VR environment, or their voices not being detected while speaking into the microphone.

Figure 4

Frequency of Coded Themes for Short-Answer Question 8

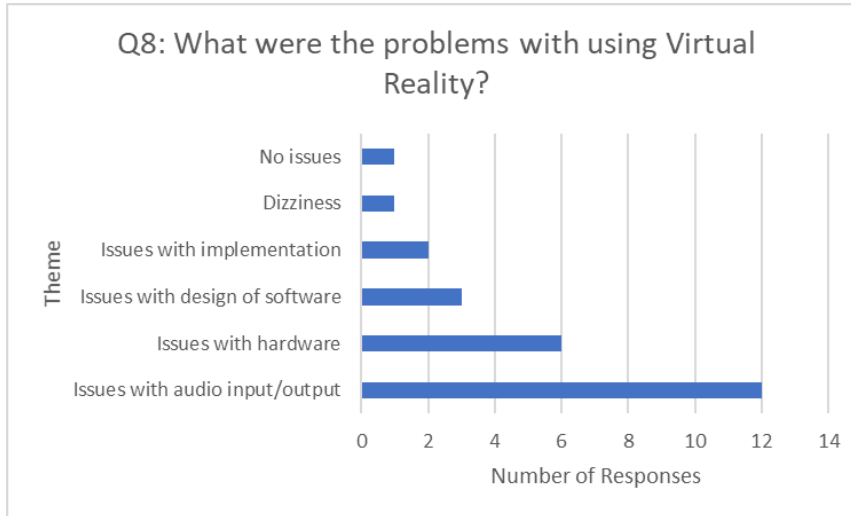
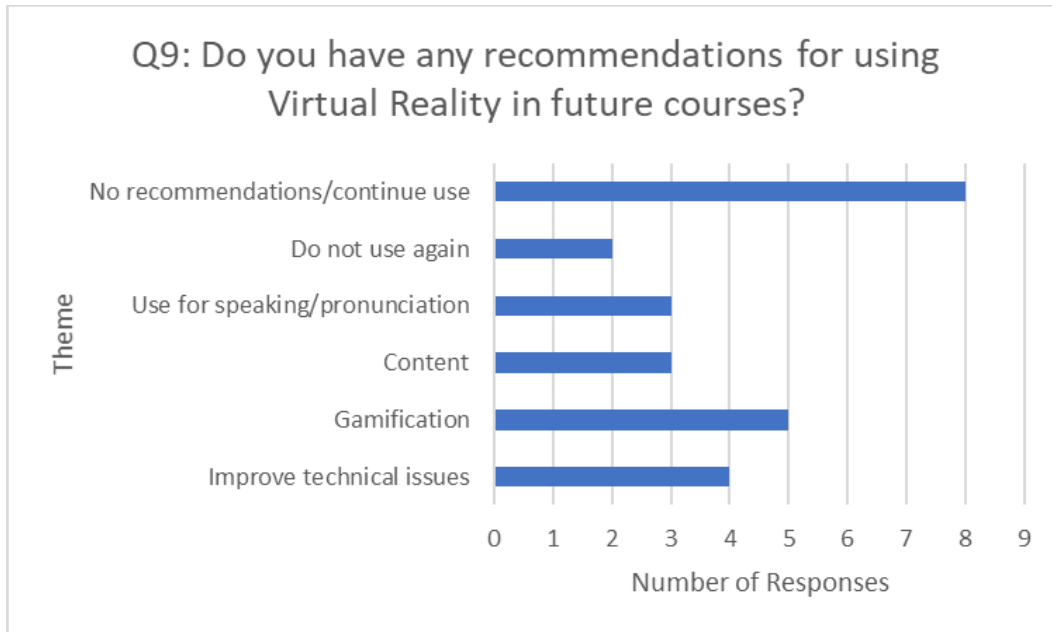


Figure 4 displays the coded themes and number of responses for each theme in Question 9. The most frequently selected theme (8 respondents) was “No Recommendations/Continue Use”. This theme emerged from comments respondents made about not having any recommendations or only recommending that the institution continue to use VR.

Figure 5

Frequency of Coded Themes for Short-Answer Question 9



4.3.1 Summary of Quantitative Results

The results for research questions one and two were obtained via statistical analysis, where interrater reliability and descriptive statistics were calculated and inferential statistics were obtained from the Wilcoxon signed-rank and the Mann-Whitney U tests. Participants in the experimental group showed significant pre-post differences in Tasks 1, 2, and 3 for appropriateness, and Tasks 2, 3, and 4 in grammar, while the control group had only one instance of significant pre-post difference (Task 3 appropriateness). When comparing gains in improvement for pragmatic performance between both groups, the control group did not show any significant gains for any of the tasks, while the experimental group showed significant gains in Task 1 for appropriateness and grammar, and in Task 4 for grammar, making Task 1 the task with

the most gains. These quantitative findings are positive and in favor of the experimental group, supporting the hypothesis that learners who practice speech act sequences (as role-plays) in an immersive VR environment where they receive individual practice and feedback show higher gains in learning than their counterparts who practice role-plays in a traditional classroom environment.

4.3.2 Summary of qualitative results

The results for research question three were obtained by administering an attitudinal questionnaire to the experimental group participants. The majority of participants responded positively when asked about their experience using VR for language learning in this study, expressing strong agreement with VR-positive statements and answering the open-ended questions with positive commentary pertaining to themes such as novelty and practice. When asked about the challenges of using VR in class, many participants commented that the audio input and output did not work at times, causing delays in practice. In spite of this technical difficulty, most participants reflected positively about the affordances of immersive VR and their experience in the study. Overall, the quantitative and qualitative findings shown in this chapter are in favor of using HiVR for L2 pragmatics learning in a classroom setting and beyond.

CHAPTER 5

DISCUSSION AND CONCLUSION

The present study examined the effects of L2 pragmatics instruction and practice in a HiVR environment on IEP students' pragmatic performance in distinct speech act sequences (such as request, refusal, and apology), as well as their attitudes and perspectives on using VR in an ESL classroom. This chapter discusses the results of this study which answered the three research questions regarding pragmatic performance as a result of role-play practice with and without VR. After discussing the results to the research questions (sections 5.1 -5.3), the limitations and pedagogical implications are mentioned (sections 5.4-5.5), the method of evaluation (5.6) followed by future research recommendations and conclusions (5.7-5.8). The procedure of the study will be reviewed in the following sections, along with the findings for each research question.

5.1 Research Question One

The purpose of this research question was to explore the usability of the VR environment for pragmatics role-play practice to see if it would help learners improve their pragmatic performance in face-to-face conversations with native speakers. The findings from the Wilcoxon signed-ranks test showed that the experimental group demonstrated substantial improvement (significant pre-post differences) in pragmatic performance for the majority of tasks (1,2,3 for appropriateness and 2, 3, and 4 for grammar), while the control showed significant improvement for only one measure of one task. Of all tasks, the findings related to Task 3 were the most intriguing. Among the three significant values found for the experimental group in the appropriateness measure,

the most significant of them was that of Task 3. Interestingly, the only instance in which the control group showed a significant pre-post difference was in the appropriateness measure of Task 3, and the significance value was slightly stronger than that of the experimental group. Possible reasons for this are that the majority of participants claimed to like Task 3 (Asking for a Favor) the most (see RQ3 results) and due to liking it, they performed well on it. Another reason could be that participants had not yet learned the conditional tenses and so they did in Task 3, which affected their pre-post grammar score.

Generally, finding for research question one confirm existing research in L2 pragmatics instruction, as they show that L2 pragmatics can be learned and it is teachable (Kasper & Rose, 1999), and that role-play activities are beneficial to pragmatics learning as they have commonly served as the productive practice aspect in explicit pragmatics instruction, which is the type of instruction that shows great learning effects (Taguchi, 2015; Nguyen, 2019; Félix-Brasdefer, 2010; Jeon & Kaya, 2006; Plonsky & Zhuang, 2019). Theoretical support for these findings is also present, as positive gains in learning were the result of explicit instruction that raised awareness of the target form (Schmidt, 2001), and required participants to complete productive practice activities (role-plays) to aid learning target forms through practice (Swain, 1995; Ellis 2005, Long 1985). Hence, it is apparent that practicing role-plays in a VR environment is at least as good as role-play practice in a traditional classroom environment because the results for this research question showed only positive effects for VR role-play practice.

5.2 Research Question Two

Research question two addressed the more crucial question at hand, which compares gains in learning between the experimental and control groups to see if learning L2 pragmatics in a HiVR environment, where learners received individual practice and feedback for role-play tasks, would result in higher gains in learning than practicing role-play tasks in a traditional classroom setting (where practice was limited and feedback was absent). The Mann Whitney U test revealed positive findings in support of the experimental group for at least two tasks in each measure. For the appropriateness measure, significant gains were observed for the experimental group over the control group for Tasks 1, 2, and 4, and for grammar in Task 1 and 3. The most significant gains score was observed in Task 1 for appropriateness. Within these findings, Task 1 was the only task in which participants showed significant gains in both measures.

First, Tasks 2 and 4, participants in the experimental group showed substantial gains in appropriateness, but not grammar. This is most likely due to the fact that in Task 2 (Apologizing for Being Late), participants benefitted from practicing this role-play and learned to use the conversational moves that are expected when one arrives late, and the cultural context in which this occurs, so improvement in grammar was unlikely in this task. Likewise, Task 4 (Asking about Private Life Politely) grammar presented in the role-play scripts only used present and past tenses, which participants had learned in previous courses, making it less cognitively demanding (grammar-wise, not appropriateness-wise) than other tasks.

The experimental group outperformed the control group in grammar for Tasks 1 and 3, and these are tasks that are more grammatically complex, as they use the conditional tenses in the role-play, along with the present, past, and future tenses. Conditional tenses are taught at the intermediate level in the IEP, so it is likely that the grammar was new to participants, therefore they showed higher pre-posttest improvement. Another reason for having succeeded in Task 3 (Asking for a Favor), specifically, (grammar), is that in the attitudinal questionnaire, the majority of participants (in the experimental group) claimed that Task 3 was their favorite task. Due to participants liking the task, they may have performed better; however, it is unknown whether the control group also liked this task the most, and the reasons for liking the task are unknown. It is likely that Task 3 was also favored by the control group as Task 3 grammar was the only task where participants in the control group made nearly significant pre-post improvement. Likewise, it is possible that participants liked this task because they were already familiar with it and they found it useful.

As for Task 1, Task 1 was unique in that it was the only task in which there were significant gains in both measures. Possible reasons for higher gains in learning Task 1 (Declining an Invitation) for both measures are various. When designing this study, I did not sequence the task lessons according to any specific criteria, hence, the tasks were not presented in order of imposition, complexity, nor difficulty (or by any other characteristic) so it is entirely possible that Task 1 was more complex or difficult than other tasks and therefore participants showed the highest gains in pre-post improvement for this task. Declining an Invitation (Task 1) was the only refusal speech act sequence in

this study; all other speech act sequences were either requests (Tasks 3 and 4) or an apology (Task 2), making Task 1 stand out as the only task of its type. It is possible that participants were already familiar or more comfortable with Tasks 2, 3 and 4, especially Tasks 2 and 3, as it is common for participants in this context to apologize for being late and ask for favors; being as I am familiar with the context of the participants (and form part of it), I have observed that students in the IEP frequently apologize for being late (especially to class) and ask for favors from classmates and teachers and I can attest to their familiarity with the tasks.

The primary reason one might give in support of participant success in learning how to carry out Task 1 over other tasks is the fact that learners were less familiar with the appropriateness and grammar required of this task prior to learning it, and that the gains in learning are largely due to pre-post differences. Apart from familiarity, it is possible that because Task 1 was the first treatment, participants may have been more excited and motivated in their learning, which led to higher gains in comparison to other tasks (as observed in Kasper, 1997). More importantly, the experimental group's superior gains for Task 1 (over other tasks) is most likely due to differences in task complexity (Robinson, 2007, 2015). The cognitive demands of Task 1 were greater than those of other tasks because the speech act sequence of Task 1 required not only a refusal, but an apology, an excuse/justification, and a suggestion. Tasks 2, 3, and 4 were arguably less complex as they required fewer conversational moves and were therefore less cognitively demanding. As Task 1 was more complex than the other tasks, participants needed maximized instruction, practice, and feedback in order to learn it (as these three elements

are necessary for effective explicit pragmatics instruction - as seen in Plonsky & Zhuang, 2019). The result of having successfully learned something that was initially less familiar to participants most likely created a larger pre-posttest difference, resulting in Task 1 as the task in which participants improved the most.

The aforementioned reasons help explain why Task 1 was learned more successfully than other tasks, however, they do not explain why the experimental group outperformed the control group in Task 1 (and other tasks). Both control and experimental groups received instruction, however, the experimental group received significantly more practice and feedback than the control group given the affordances of the VR environment. While the explanation portion of the explicit instruction was most likely helpful, it is apparent that the enhanced practice and feedback are what caused participants in the experimental group to perform substantially better than their counterparts in the control group. This is not surprising given what is known from previous studies that have used digital tools for oral tasks, which reiterate the importance of the digital tool providing feedback to learners in order to be effective and for teachers to view the tool as worthy of using (Timpe-Laughlin et al., 2022).

Indeed, the amount of enhanced individual practice and feedback participants received in the VR environment is not possible in a traditional classroom environment; this instructional condition explains why the experimental group outperformed the control group in the majority of tasks and measures (not only Task 1). Certainly, the continuous, individualized practice and feedback are the unique affordances VR has to

offer and should be the primary motivation behind using VR in education settings such as the one in this study.

In previous research studies that compared EFL learning effects in VR versus non-VR contexts, such as those mentioned in Qiu et al.'s (2023) meta-analysis, larger effects were observed for VR treatment groups. Though none of the studies in Qiu et al.'s (2023) meta-analysis measured pragmatics specifically, the fact that learning effects were observed for other language skills led me to assume that L2 pragmatics would follow suit. More specifically, I hypothesized that higher gains in learning would be achieved as a result of using HiVR. Though Qiu et al.'s (2023) meta-analysis featured a mix of LiVR and HiVR environments, the two types of VR were not compared; however, I suspected that if studies with LiVR could render positive learning effects, the HiVR would yield even greater ones because it affords learners a simulated, embodied experience that is closer to real life (Kaplan-Rakowski & Gruber, 2021; Gall et al. 2021), learners could interact with someone of a higher level of proficiency (in this case, a pre-recorded video of someone) (Long, 1981), and the quantity of practice and feedback (Long, 1985) would be more than that of the traditional classroom model, causing the VR group to outperform the non-VR group. Overall, the results of this study confirmed my hypothesis, showing that participants who use VR for practice and feedback in a classroom environment will demonstrate superior pragmatic performance than those who practice in a face-to-face modality only.

5.3 Research Question Three

Research question three was more exploratory than research questions one and two. Research question three was concerned with discovering the experimental groups' attitudes and perceptions towards using HiVR for language learning purposes in a classroom environment. Generally, there was a mix of positive and negative responses regarding the use of VR and the software itself, but overall, most participants were in favor of using VR and the reported challenges were mostly about technical issues. The responses from the open-ended questions reiterate previously existing research on learner attitudes and perceptions of VR; learners like VR because it is novelty, (exciting, fun, motivating, new, engaging) as observed in Soto et al.'s (2020) study that used ImmerseMe for EFL instruction and other studies that reference learning with VR (Bae et.al, 2017; Peterson, 2018), it helps reduce foreign language anxiety, as seen in Thrasher, (2022), and that it simulates real-life practice as described in Gall et al. (2021) and Kaplan-Rakowski and Gruber (2021, 2023).

Specific findings from the attitudinal questionnaire not only confirm previously existing research, they support findings from the quantitative section of the current study. The independent variables of this study were modality, practice and feedback, and the results of research questions one and two showed strong effects for learning speech act sequences in a HiVR environment where participants received individual practice and feedback. Hence, the affordances that are unique to HiVR (individualized practice and feedback), affected learning outcomes positively. Responses from the attitudinal questionnaire support this observation. For example, when participants were asked what

their favorite thing about learning VR was (in the open-ended section of the survey), 20% of respondents commented that the practice and/or feedback features of the VR were what they liked the most. I assume that liking these features also means that participants found them to be helpful. This reinforces the idea that practice and feedback in the HiVR environment were what helped the participants in the experimental group outperform their counterparts in the control group.

Another finding from the attitudinal questionnaire that supports the findings in the quantitative section is that of Question 5. When participants were asked to select the VR lesson they liked the most, Task 3 (Asking for a Favor) had the highest number of responses of the four tasks, accounting for 48% of the respondents. This qualitative finding could help explain why participants in both groups performed so well on this particular task, according to the Wilcoxon signed-ranks test for research question one.

It is quite possible that they showed significant pre-post learning outcomes because they liked the task. Deci and Ryan's (2000) Self-Determination theory suggests personal interest and enjoyment surrounding a particular topic can lead to enhanced learning outcomes. In addition, it could be observed that Task 3, "Asking for a Favor" is a request that is likely to be less face-threatening than the other request sequence used in this study, which was Task 4 (Asking about Private Life Politely), and due to this, it was easier for participants to learn. This observation is consistent with previous L2 pragmatics research that observes that learners are more likely to have better learning outcomes for tasks that are less face-threatening and therefore low-anxiety (Bardovi-Harlig, & Hartford, 1990, Kasper, 1997).

The challenges reported with using VR were mostly technical, such as audio input/output issues; participants reported that at times, the software did not recognize their utterances, and they occasionally could not hear the responses of the person speaking to them in the VR environment. Few students reported issues with the HMDs, saying the headsets were not comfortable to wear. Indeed, technical issues can often interfere with learning outcomes in CALL-based environments (Warschauer, 2004), therefore it is imperative that these issues be addressed if a new technology such as HiVR is to be used in a classroom environment.

5.4 Pedagogical Implications

This dissertation study supports the use of HiVR for L2 pragmatics learning in the US university IEP classroom context. The results from the data analysis imply that VR can be used for pragmatically-based role-play practice and that the affordances of VR-based practice are superior to traditional classroom-based role-play practice in that they give each student immediate, individualized and continuous practice and feedback, which is not possible nor practical in a traditional classroom environment. Indeed, students and teachers can benefit from using VR in an IEP classroom context, especially for L2 pragmatics, an aspect of language acquisition that has traditionally been ignored in comparison to other skills (such as reading, writing, listening and speaking, grammar, etc.). Previous research strongly suggests that L2 pragmatics can be learned via explicit instruction, showing preference for explicit over implicit (Plonsky & Zhuang, 2019, Norris & Ortega, 2001; Taguchi, 2015; Taguchi, 2011; Takahashi, 2010b; Jeon & Kaya,

2006; Rose, 2005), and that role-play practice is an effective way for learners to develop pragmatic competence in the target language (Félix-Brasdefer, 2010).

In terms of implementation, HiVR requires high-priced equipment (HMDs and hand controllers) and software which is a barrier for many schools, institutions, and populations of students. Prices for HMDs range anywhere from \$200-\$2,000 USD per set, and individual subscriptions to off the shelf language learning software like ImmerseMe are also priced relatively highly at \$300 USD per year (*Pricing — ImmerseMe - Using Virtual Reality to Learn Languages*, n.d.), per person. In addition, high-speed internet access is required for these products to work properly, and loss in internet connectivity or slow internet speeds will prevent users from accessing the lessons. It is also recommended that HiVR be used in a setting where there are technical support personnel on staff to maintain the equipment and help users troubleshoot issues.

Last, it is important to teach learners how to use the equipment and how to troubleshoot issues that require rebooting the HMDs or logging in and out of the software site where they access the lessons. During the treatment sessions in this dissertation study, four to five lab assistants were present at each session, and they constantly helped participants troubleshoot technical issues, use the HMDs and hand controllers properly, and navigate ImmerseMe. For example, in this study, participants frequently moved out of range and were automatically logged out of the equipment and platform, and the lab assistants had to help them log back in. It is also imperative that students be afforded the opportunity to become comfortable with the VR hardware and navigate the platforms or

software websites on their own. Instructors should not assume that all of their students already know how to use HiVR.

Before starting the treatment of this study, participants in the experimental VR group were given an introductory session where they practiced navigating the ImmerseMe platform, and this was helpful and time-saving for the following treatment sessions, however, it was my observation that one practice session was not enough for all participants to become accustomed to using HiVR, as I observed many of them needing help with navigation during the treatment session. Therefore, my recommendation for teachers and schools who would like to adopt HiVR in their classrooms is that they ensure access to high-speed internet, funding for equipment and software, on-demand technical support, and practice sessions that allow students to navigate freely without having to complete graded activities.

Regarding instructional integration, I do not recommend removing the teacher from the classroom nor relying solely on the lessons/practice activities on the VR platform (ImmerseMe, in the case of this study), especially in the case of L2 pragmatics learning. Learners need explicit instruction in order to raise their metapragmatic awareness (Bardovi-Harlig & Mahan-Taylor, 2003), otherwise they would only complete the role-play practices which do not guarantee that they would recognize and process conversational moves and expected interaction sequences, nor the sociocultural implications for the target speech act sequences (Schmidt, 1993). Existing research provides strong evidence for explicit instruction over implicit learning of L2 pragmatics even in study abroad contexts such as the one in this study (Alcón-Soler, 2015; Winke &

Teng, 2010; Halenko & Jones, 2011; Bouton; 1999). ImmerseMe does not include instruction nor explanation in the lessons they provide, but rather, practice that appears to be task-based, which is very useful for in-class or for practice outside of the classroom.

5.5 Limitations

Limitations of the current study are various: the sample size, institutional constraints, context/setting, age of participants, method of testing, the rating of tests, and technical issues. First, the sample size of participants in this dissertation study is not large, with 34 participants in the experimental group and 43 in the control group. However, my sample size was comparatively larger than other sample sizes in previous VRALL studies which were between 14-34 (Qiu et al., 2023; Dhimolea et al., 2022; Parmaxi, 2020). The modest sample of this study is due mostly to institutional constraints related to time, cost, and logistics. I recruited participants from the course level with the highest enrollment (intermediate) during the semester with the highest enrollment (Fall), in order to maximize participation, and the recruitment yielded 77 participants out of 88 students. The size of the IEP itself is moderate, therefore the number of possible participants was limited.

I carried out the study during the first half of the Fall semester, which is the IEP program's first eight-week session of the academic year. As the treatment for the study was instructional, treatment sessions needed to align with existing classes and class times. Likewise, I had to work with IEP staff to adapt parts of the curriculum for the communication course to accommodate my study (which interrupted regular procedures

and processes in the IEP). In addition, securing five weeks of lab sessions with all necessary equipment and technical support staff at my institution's VR lab was costly.

5.6 Participants and setting

The target second language, context of the study, and age of participants, perpetuate existing VRALL research as being English-centered with participant populations being mostly university students who are learning English as a foreign language or native English speakers who are learning a foreign language, as seen in reviews from Qiu et al., (2023), Dhimolea et al., (2022), and Parmaxi (2020). The current study did not address the need for more VRALL research with participants who are of diverse age groups (age groups other than 18-25 years old) in diverse settings (contexts other than university), and are not speakers or learners of English. I was not able to address this limitation in my dissertation study because VRALL for ESL/EFL is my primary concern and interest as a TESOL practitioner and therefore English was the target language of study, and my place of work (a university IEP) is my primary context of concern. Learners of diverse age groups, contexts, and backgrounds may have completely different experiences with VR than the participants of this study, and future research should be conducted to meet this need.

5.6.1 Method of Evaluation

Participants were administered pre and posttests that were open role-plays. Though this method has been used frequently in existing L2 pragmatics research and is thought to be valid, it does have limitations (Félix-Brasdefer, 2010). One major limitation is that role-plays are hypothetical, so they simulate a hypothesized reality which is not the

same as reality itself. For this reason, L2 pragmatics methods have evolved to using Conversation Analysis to study pragmatic competence, as recommended by Kasper (2006). In a role-play, many paralinguistic factors are not accounted for, such as gender, socioeconomic or cultural background, real-life context, relationship between interlocutors, etc. In the current study, it was not logistically possible for me to control for gender or socioeconomic and cultural background due to time constraints and number of native speaker domestic students per international student participant. By the end of each treatment session, we were left with only 10-15 minutes to complete the role-play posttests and these needed to be done fairly quickly in order to ensure all participants were role-played. Participants had to spontaneously pair up with whichever domestic student was available to pair with them at the time.

Last, one important characteristic of the posttests was that they were immediate; it is unknown if participants were able to retain the pragmatic knowledge acquired during instruction and practice beyond the same day of treatment. In future research, it may be beneficial for researchers to give delayed posttests to measure the permeability of instructional treatment. In addition to delayed posttests, future qualitative studies could benefit from surveying or interviewing participants after each treatment to learn about their perceived interactions from a first-person perspective. Surveying the domestic students who administered the role-play tests could also provide insight into perceptions of their interactions with international students.

5.6.2 Rating

Though there are several rater-characteristics that make the raters reliable and qualified to rate tests for this study, there were some limitations in rating: possible researcher bias and the rubrics used for rating. Due to time, cost, and institutional constraints, it was not feasible for me to recruit two raters for this study; I was only able to recruit one, as there were hundreds of tests to rate which required several hours of someone's time. Because of this, I had to serve as the second rater. This circumstance created the possibility of researcher bias in rating. In order to minimize this risk, I made sure to label students with identification numbers that were random and had no indication of their group (experimental or control), so that when I was rating the tests I could not immediately recall which group they belonged to. Also, because I had to edit all audio recordings before sharing them with the other rater, I did not hear students identify themselves by name so I could not immediately know which section they were in. The labeling of test recordings and editing of test audios was done prior to rating the tests.

The other possible limitation is the rubric for appropriateness and grammar. Though I adapted these rubrics from Taguchi (2012) and they have been used in other studies in which recorded speech is evaluated for pragmatic competence or performance, they are blunt instruments with a 0-5 point rating scale, and with such instrument it is difficult to measure improvement in a short amount of time, as learners would have to make significant progress to move from one score to the next (i.e. 3 to 4, or 2 to 5).

Also, Taguchi's (2012) framework for evaluating pragmatic competence included a measurement of fluency, in addition to a 5-point scale rubric for appropriateness and for

grammar. I did not measure participant fluency separately nor did I calculate how many words per minute they spoke. Instead, I included fluency in parts of the descriptive rubrics for appropriateness and grammar (fluency affects both of these measures). Also, I did not calculate words per minute as benchmarks for words per minute are typically set by native speaker standards and this could be an unrealistic benchmark. Nonetheless, it may have been beneficial to evaluate participants' fluency separately, as Taguchi (2012) recommended.

5.6.3 Technical issues

Another potential limitation for the current study is that of technical issues. The results could have been affected by the technical issues experienced by experimental group participants during treatment sessions. There were instances in which the VR equipment would freeze, log students out, shut off, or fail to recognize audio input. The lab assistants were present during treatment sessions to help participants as they encountered issues, however, troubleshooting took time away from potential practice. Likewise, most participants were not experienced users of HiVR, so they did not know how to troubleshoot on their own. More experience, comfort, and confidence using HiVR may have led to more pronounced gains. Existing research in CALL (Hubbard, 2008; Levy & Stockwell, 2006; Warschauer, 2004) affirms that technical issues do indeed interfere with learning in a classroom setting and this can cause frustration to learners.

5.7 Recommendations for Future Research

VRALL, especially HiVRALL, research is relatively new and there is comparatively little existing research in this area. Very few existing studies show effects

with statistical significance, and large sample sizes are scarce. HiVRALL needs to be researched in contexts where English is not the L1 of the participants nor the target language, and with age groups other than university students (18-25 years old). Studies with longer treatment times (more than one semester) are needed in order to test the effectiveness of VR in helping learners retain knowledge and pedagogical sustainability of VR integration. In order for more HiVR studies to take place, HMDs and off-the-shelf VR software must become more readily available at accessible price points. Otherwise, HiVRALL research will only continue to occur in high-access environments with elite learners. Existing research is highly focused on vocabulary learning and some on oral communication (Dhimolea et al., 2022), however, more studies on HiVR for L2 pragmatics (such as this dissertation study) are needed. Within L2 pragmatics, it would be helpful if there were more studies that compared learning effects for using VR to learn different speech act sequences. Likewise, methods for evaluating learner pragmatic performance within the VR environment would be beneficial for future research.

5.8 Conclusions

Overall, the findings in this study were positive and in support of using HiVR for L2 pragmatics learning. Participants in the experimental group (VR) in this dissertation study showed significant pre-posttest improvement in all tasks during the instructional treatment, and when compared to the control (non-VR) group, they showed higher gains in learning in three of the four tasks. In their majority, participants commented positively about their experience using VR for learning English L2 pragmatics. These findings, both quantitative and qualitative, reiterate those of current, empirical research for learning

other foreign language skills (vocabulary, oral communication, pronunciation, etc.) in LiVR and HiVR environments; existing research shows larger effects for learning with VR versus traditional classroom practice, and positive perceptions and attitudes from learners.

It is highly probable that research concerning L2 pragmatics instruction in CALL environments will continue to grow with time as technology advances and technological tools such as HiVR become more ubiquitous. Though this dissertation study was the first to examine the learning effects of L2 pragmatics instruction using HiVR in an IEP educational context, it is unlikely to be the last. As learners in study abroad and foreign language contexts become increasingly more aware of their need to develop pragmatic competence in order to interact appropriately and successfully with interlocutors in their contexts, they are bound to seek instruction and digital tools to help them meet their goals. For this reason, practitioners, researchers, and educational institutions ought to consider piloting a VR-based instructional treatment in their contexts, as it could benefit learners, especially in the area of L2 pragmatics.

REFERENCES

- Akiyama, Y., & Cunningham, D. J. (2017). Synthesizing the Practice of SCMC-based Telecollaboration: A Scoping Review. *CALICO Journal*, 35(1), 49–76. <https://doi.org/10.1558/cj.33156>
- Alcón-Soler, E. (2007). Fostering EFL learners' awareness of requesting through explicit and implicit consciousness-raising tasks. In García Mayo (ed.), *Investigating tasks in formal language learning* (pp 221–241). Clevedon: Multilingual Matters.
- Alcón-Soler, E. (2015). Pragmatic learning and study abroad: Effects of instruction and length of stay. *System*, 48, 62–74.
- Alcón-Soler, E. (2018). Effects of task-supported language teaching on learning how to mitigate email requests. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics*. John Benjamins.
- Austin, J. L. (1962). *How to do things with words*. J.O. Urmson & Marina Sbisa (Eds.). Harvard University Press.
- Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford, England; Oxford University Press.
- Bae, H., Kang, M., Kim, B., & Lee, H. (2017). Effectiveness of Virtual Reality for Learning Based on User Experience. *Educational Technology & Society*, 20(4), 133–140.
- Bardovi-Harlig, K. (2001). Evaluating the empirical evidence: Grounds for instruction in pragmatics? In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 13–32). Cambridge University Press.
- Bardovi-Harlig, K. (2003). Understanding the role of grammar in the acquisition of L2 pragmatics. In A. Martínez Flor, E. Usó Juan, & A. Fernández Guerra (Eds.), *Pragmatic competence and foreign language teaching* (pp. 25–44). Publicacions de la Universitat Jaume I.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(1), 68–86.
- Bardovi-Harlig, K. (2017). Acquisition of L2 pragmatics. *The Routledge handbook of instructed second language acquisition*, 224-245.
- Bardovi-Harlig, K., & Hartford, B. S. (1990). Congruence in native and nonnative conversations: Status balance in the academic advising session. *Language learning*, 40(4), 467-501.

- Bardovi-Harlig, K., & Mahan-Taylor, R. (2003). Teaching pragmatics. Washington DC: U.S. Department of State, Office of English Language Programs.
- Belz, J. A., & Kinginger, C. (2002). The cross-linguistic development of address form use in telecollaborative language learning: Two case studies. *The Canadian Modern Language Review*, 59(2), 189–214.
- Belz, J. A., & Kinginger, C. (2003). Discourse options and the development of pragmatic competence by classroom learners of German: The case of address forms. *Language Learning*, 53(4), 591.
- Biesenbach-Lucas, S. (2006). Making requests in e-mail: Do cyber-consultations entail directness? Toward conventions in a new medium. In K. Bardovi-Harlig, J. C. Félix-Brasdefer, & A. S. Omar (Eds.), *Pragmatics & language learning* (Vol. 11, pp. 81–108). National Foreign Language Resource Center, University of Hawai'i at Mānoa.
- Biesenbach-Lucas, S. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning & Technology*, 11(2), 59–81.
- Bouton, L. F. (1999). Developing nonnative speaker skills in interpreting conversational implicatures in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47–70). Cambridge: Cambridge University Press.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- Chang, Y. Y., & Hsu, Y.-P. (1998). Requests on email: A cross-cultural comparison. *RELC Journal*, 29(2), 121–151.
- Chen, C. F. E. (2006). The development of e-mail literacy: From writing to peers to writing to authority figures. *Language Learning & Technology*, 10(2), 35–55. <https://www.lltjournal.org/item/2547>
- Chen, Y. (2015). Developing Chinese EFL learners' email literacy through requests to faculty. *Journal of Pragmatics*, 75, 131-149.

- Codina-Espurz, V., & Salazar-Campillo, P. (2019). Openings and closing in emails by CLIL students: A pedagogical proposal. *English Language Teaching*, 12(2), 57.
- Cunningham, D. J. (2016). Request modification in synchronous computer-mediated communication: The role of focused instruction. *The Modern Language Journal*, 100(2), 484–507.
- Cunningham, D. J. (2017). Second language pragmatic appropriateness in telecollaboration: The influence of discourse management and grammaticality. *System*, 64, 46–57.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.
- Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, 66(5), 810–824.
- DuFon, M. A. (1999). *The acquisition of linguistic politeness in Indonesian as a second language by sojourners in naturalistic interactions* (Doctoral dissertation, University of Hawai'i at Manoa).
- Dombi, J. (2020). Email requests in an ELF academic setting: “Good luck to you, Doctor Judit...” *Applied Pragmatics*, 2(1), 54–79.
- Ekiert, E., Lampropoulou, S., Revesz, A., & Torgersen, E. (2018). Chapter 10: The effects of task type and L2 proficiency on discourse appropriacy in oral task performance. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics* (pp. 247-264). John Benjamins Publishing Company.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141–172.
- Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 313-348). Routledge.
- Félix-Brasdefer, J. C. (2008). Pedagogical intervention and the development of pragmatic competence in learning Spanish as a foreign language. *Issues in Applied Linguistics* 16, 49–84.

- Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance: DCTs, roleplays, and verbal reports. In A. Martínez Flor & E. Usó Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 41–56). John Benjamins Pub. Company.
- Félix-Brasdefer, J.C. & Bardovi-Harlig, K. (2010). “I’m sorry. Can I think about it?” The negotiation of refusals in academic and non-academic contexts. In D. Taksuki & N. Houck (Eds.), *Pragmatics from research to practice: Teaching speech acts* (pp. 163-180). Alexandria VA: TESOL.
- Félix-Brasdefer, C. (2012). E-mail requests to faculty: E-politeness and internal modification. In M. Economidou-Kogetsidis & H. Woodfield (Eds.), *Interlanguage request modification* (pp. 87–118). John Benjamins Pub Company.
- Gall, D., Roth, D., Stauffert, J. P., Zarges, J., & Latoschik, M. E. (2021). Embodiment in virtual reality intensifies emotional responses to virtual stimuli. *Frontiers in Psychology*, 12.
- García-Gómez, A. (2020). Intercultural and interpersonal communication failures: Analyzing hostile interactions among British and Spanish university students on WhatsApp. *Intercultural Pragmatics*, 17(1), 27–51.
- Gass, S. M. & N. Houck. (1999). *Interlanguage Refusals: A Cross-Cultural Study of Japanese-English*. Nerk: Mouton de Gruyter.
- González-Lloret, M. (2008). Computer-mediated Learning of L2 Pragmatics. In E. A. Soler & A. Martinez-Flor (Eds.), *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing* (pp. 114–132). Multilingual Matters.
- Gonzales, A. (2013). Development of politeness strategies in participatory online environments: A case study. In N. Taguchi & J. M. Sykes (Eds.), *Language Learning & Language Teaching* (Vol. 36, pp. 101–120). John Benjamins Publishing Company.
- González-Lloret, M. (2019). Technology and L2 pragmatics learning. *Annual Review of Applied Linguistics*, 39, 113–127.
- González-Lloret, M. (2021). L2 pragmatics and CALL. *Language Learning & Technology*, 25(3), 90–105. <http://hdl.handle.net/10125/73451>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41-58). Academic Press.

- Halenko, N., & Jones, C. (2011). Teaching pragmatic awareness of spoken requests to Chinese EAP learners in the UK: Is explicit instruction effective? *System*, 39(2), 240–250.
- Han, S. (2005). The interlanguage pragmatic development of the speech act of requests by Korean non-native speakers of English in an ESL setting [Doctoral Dissertation, University of Pennsylvania]. ProQuest Dissertations and Theses Global.
- Hassall, T. (2013). Pragmatic development during short-term study abroad: The case of address terms in Indonesian. *Journal of Pragmatics*, 55, 1–17.
- Holden, C., & Sykes, J. M. (2012). Mentira: Prototyping language-based locative gameplay. In S. Dijkers, J. Martin, & B. Coulter (Eds.), *Mobile media learning: Amazing uses of mobile devices for teaching and learning* (pp. 111–131). Pittsburgh, PA: ETC Press.
- House, J. (1996). Developing pragmatic fluency in English as a foreign language. *Studies in Second Language Acquisition* 18, 225–252.
- Hubbard, P. (2008). CALL and the future of language teacher education. *CALICO Journal*, 25(2), 175-188.
- Hymes, D. (1962). The Ethnography of Speaking. In T. Gladwin & W. C. Sturtevant (Eds.), *Anthropology and human behavior* (pp. 13–53). Anthropological Society of Washington.
- Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz, J. J. & Hymes, D. (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35-71). New York: Holts Rinehart & Winston.
- Ishihara, N. (2019). Identity and agency in L2 pragmatics. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 161–175). Routledge.
- Ishihara, N., & Tarone, E. (2009). Subjectivity and pragmatic choice in L2 Japanese: Emulating and resisting pragmatic norms. In N. Taguchi (Ed.), *Pragmatic competence* (Book in Pragmatics folder; pp. 101–128). Mouton de Gruyter.
- Iino, M. (2006). Norms of interaction in a Japanese homestay setting: Toward a two-way flow of linguistic and cultural resources. In M. A. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 151-174). Blue Ridge Summit: Multilingual Matters.
- ImmerseMe. (n.d.). *Pricing*. <https://immerseme.co/pricing>

- Jeon, E. H. & T. Kaya (2006). Effects of L2 instruction on interlanguage pragmatic development. In N. John & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165-211). Philadelphia: John Benjamins Pub. Company.
- Kaplan-Rakowski, R., & Gruber, A. (2021). One-on-one foreign language speaking practice in high-immersion virtual reality. In C.C. Cheng, K.E. Chang, Y.T. Sung, & P. Ling (Eds.), *Chinese language learning sciences* (pp. 187–202).
- Kaplan-Rakowski, R., & Gruber, A. (2023). The impact of high-immersion virtual reality on foreign language anxiety. *Smart learning environments*, 10(1), 46.
- Kasper, G. & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition* 13(2): 215–247.
- Kasper, G. (1997). Can pragmatic competence be taught? In L. F. Bouton (Ed.), *Pragmatics and language learning* (Vol. 8, pp. 238-259). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- Kasper, G., & Rose, K. R. (Eds.). (1999). *Pragmatics and language teaching*. Cambridge University Press.
- Kasper, G. (2001). Teaching pragmatics in language classes. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 135–157). Cambridge University Press.
- Kasper, G., & Rose, K. (2002). Chapter 2: Theories of second language pragmatic development. In *Pragmatic development in a second language* (pp.13-61). Oxford: Blackwell.
- Kasper, G. (2006). Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 317-334). Lawrence Erlbaum Associates.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. New York: Palgrave Macmillan.

- Leech, G. (1983). *Principles of pragmatics*. London: Longman.
- Lan, Y. J. (2020). Immersion, interaction, and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology*, 24, 1–5. <http://hdl.handle.net/10125/44704>
- Levy, M., & Stockwell, G. (2006). *CALL Dimensions: Options and issues in computer-assisted language learning*. Routledge.
- Li, S., Taguchi, N., & Xiao, F. (2019). Variations in rating scale functioning in assessing pragmatic performance in L2 Chinese. *Language Assessment Quarterly* 16(3): 271–293.
- Lin, T. J., & Lan, Y. J. (2015). Language learning in virtual reality environments: Past, present, and future. *Educational technology & society*, 18(4), 486–497. http://www.ifets.info/journals/18_4/37.pdf
- Long, M. H. (1981). Input, interaction, and second language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259-278). Annals of the New York Academy of Sciences.
- Long, M. H. (1985). Input and second language acquisition theory. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Newbury House.
- Merriam-Webster. (n.d.). Virtual reality. In Merriam-Webster.com dictionary. Retrieved March 4, 2024, from <https://www.merriam-webster.com/dictionary/virtual%20reality>
- Nguyen, T. T. M. (2018). Pragmatic development in the instructed context: A longitudinal investigation of L2 email requests. *Pragmatics*, 28(2), 217–252.
- Nguyen, T. T. M. (2019). Data collection methods in L2 pragmatics research: An overview. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp.195–211). Routledge.
- Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, 51(s1), 157–213.
- Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments*, 31(1), 172–184.

- Peterson, M. (2018). Virtual reality and the future of language learning. *Language Learning & Technology*, 22(2), 3–17.
- Plonsky, L., & Zhuang, J. (2019). A meta-analysis of second language pragmatics instruction. In N. Taguchi (Ed.), *Routledge handbook of second language acquisition and pragmatics* (pp. 287–307). Routledge.
- Qiu, X. B., Shan, C., Yao, J., & Fu, Q. K. (2023). The effects of virtual reality on EFL learning: A meta-analysis. *Education and Information Technologies*, 2, 1379–1405.
- Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. del P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Multilingual Matters.
- Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (Vol. 8, pp. 87–121). John Benjamins Publishing Company.
- Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33(3), 385–399.
- Salazar-Campillo, P. (2018). Student-initiated email communication: An analysis of openings and closings by Spanish EFL learners. *Sintagma: Revista de Lingüística*, 30, 81–93.
- Salazar-Campillo, P., & Codina-Espurz, V. (2018). Politeness in first and follow-up emails to faculty: Openings and closings. In A. Sánchez-Hernández & A. Herraiz-Martínez (Eds.), *Learning second language pragmatics beyond traditional contexts* (pp. 85–106). Peter Lang AG.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In Kasper, G., & Blum-Kulka, S. (Eds.), *Interlanguage pragmatics* (pp. 21–42). New York, NY: Oxford University Press.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 3–32.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

- Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1-23.
- Shively, R. L. (2011). L2 pragmatic development in study abroad: A longitudinal study of Spanish service encounters. *Journal of Pragmatics*, 43(6), 1818–1835.
- Siegal, M. (1996). The role of learner subjectivity in second language linguistic competency: Western women learning Japanese. *Applied Linguistics*, 17, 356–382.
- Soto, J. H. B., Ocampo, D. C. T., Del Carmen Beltrán Colon, L., & Oropesa, A. V. (2020). Perceptions of ImmerseMe Virtual Reality Platform to improve English communicative skills in higher education. *International Journal of Interactive Mobile Technologies*, 14(07), 4.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University
- Sykes, J. M. (2005). Synchronous CMC and pragmatic development: Effects of oral and written chat. *CALICO Journal*, 22(3), 399–431.
- Sykes, J. M. (2009). Learner requests in Spanish: Examining the potential of multiuser virtual environments for L2 pragmatic acquisition. In L. Lomicka & G. Lord (Eds.), *The second generation: Online collaboration and social networking in CALL* (pp. 199–234). Texas State University.
- Sykes, J. (2013). Multiuser virtual environments: Learner apologies in Spanish. In N. Taguchi & J. Sykes (Eds.), *Technology and interlanguage pragmatics research and teaching* (pp. 71–100). John Benjamins Pub Company.
- Takahashi, S. (2010a). The effect of pragmatic instruction on speech act performance. In Martínez-Flor, A. & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 127–144). Amsterdam/Philadelphia: John Benjamins Pub Company.
- Takahashi, S. (2010b). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Handbook of pragmatics* (pp. 391-421). Berlin: Mouton de Gruyter.
- Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310.

- Taguchi, N. (2012). Theoretical framework, research questions and methodology of the study. In N. Taguchi (Ed.). *Context, individual differences and pragmatic competence*. Multilingual Matters eBooks.
- Taguchi, N., & Sykes, J. M. (Eds.). (2013). *Technology in interlanguage pragmatics research and teaching*. John Benjamins Pub Company.
- Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, 48(1), 1–50.
- Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
- Taguchi, N. (2021). Application of immersive virtual reality (VR) to pragmatics data methods: Insights from interviews. *CALICO Journal*, 38(2), 181–201.
- Taguchi, N. (2022a). Dispreferred speech acts in virtual reality: Analysis of tone choices and hesitations. *System*, 106, 102793.
- Taguchi, N. (2022b). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, 56(1), 308–335.
- Tang, X., & Taguchi, N. (2020). Designing and using a scenario-based digital game to teach Chinese formulaic expressions. *CALICO Journal*, 37(1), 1–22.
- Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in L2 Chinese. *Modern Language Journal*, 105, 740–759.
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4, 91–112.
- Thrasher, T. (2022). Impact of virtual reality on L2 French learners' language anxiety and oral comprehensibility. *The CALICO Journal*, 39(2), 219–238.
- Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2022). Using spoken dialogue technology for L2 speaking practice: what do teachers think? *Computer Assisted Language Learning*, 35(5–6), 1194–1217.
- Usó-Juan, E. (2022). Exploring the role of strategy instruction on learners' ability to write authentic email requests to faculty. *Language Teaching Research*, 26(2), 213–237.
- Van Compernelle, R. A., Gomez-Laich, M. P., & Weber, A. (2016). Teaching L2

- Spanish Sociopragmatics Through Concepts: A Classroom-Based Study. *The Modern Language Journal*, 100(1), 341–361.
- Vidal, C. P., & Shively, R. L. (2019). L2 pragmatic development in study abroad Settings. In *the Routledge handbook of second language acquisition and pragmatics* (1st ed., pp. 355–371). Routledge.
- Vilar-Beltrán, E., & Melchor-Couto, S. (2013). Refusing in second life. *Utrecht Studies in Language and Communication*, 25, 23–40.
- Warschauer, M. (2004). Of digital divides and social multipliers: Combining language and technology for human development. In *Information and communication technologies in the teaching and learning of foreign languages: State of the art, needs and perspectives* (pp. 46-52). Moscow: UNESCO Institute for Information Technologies in Education.
- Williams, L., & van Compernelle, R. A. (2009). On versus tu and vous: Pronouns with indefinite reference in synchronous electronic French discourse. *Language Sciences*, 31(4), 409–427.
- Winans, M. D. (2020). Email requests: Politeness evaluations by instructors from diverse language backgrounds. *Language Learning & Technology*, 24(2), 104–118. <https://www.lltjournal.org/item/3152>
- Winke, P., & Teng, C. (2010). Using task-based pragmatics tutorials while studying abroad in China. *Intercultural Pragmatics*, 7(2).
- Young, R. F. (2011). Interactional competence in language learning, teaching, and testing. In *Handbook of research in second language teaching and learning* (pp. 426-443). Routledge.

APPENDIX A
IRB APPROVAL



EXEMPTION GRANTED

David Smith
 CLAS-H: English
 -
 bryansmith@asu.edu

Dear [David Smith](#):

On 8/15/2023 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Immersive Virtual Reality for L2 Pragmatics Instruction in an Intensive English Program at a US University
Investigator:	David Smith
IRB ID:	STUDY00018363
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	<ul style="list-style-type: none"> • Consent form Control Group, Category: Consent Form; • Consent Form Treatment Group, Category: Consent Form; • Instrument 1, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Instrument 2 Role-Plays Pre and posttests , Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Instrument 3 Surveys Pre and Post, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Protocol, Category: IRB Protocol; • Recruitment Email Control , Category: Recruitment Materials; • Recruitment Email Treatment, Category: Recruitment Materials;

--	--

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (1) Educational settings, (2)(ii) Tests, surveys, interviews, or observation (low risk) on 8/15/2023.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103). When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB".

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Emilia Gracia
David Smith
Emilia Gracia

APPENDIX B
ROLE-PLAY TESTS

Role-Play Pretests

Instructions: Read the following scenarios and complete the suggested task for the Global Launch student.

Pretest 1 (Declining an Invitation):

ASU Student: You invite your roommate to Thanksgiving dinner at your family's house. You really want them to come because it is a special occasion and you want your international roommate to feel welcome in the US. (You start the conversation)

Global Launch Student: Your roommate invites you to have Thanksgiving dinner at their family's house. You would like to go, but you have other plans for the weekend. You are planning to go on a trip with two other friends to Disneyland and you already paid for the trip. (ASU student starts the conversation)

Pretest 2 (Apologizing for being late):

ASU Student: You made plans with a classmate to meet at the library to study for the math final. You have been waiting in the library now for 45 minutes. Your classmate (GL student) has not responded to your texts. GL student then arrives. Respond to them accordingly. (GL student starts the conversation)

Global Launch Student: You made plans with a classmate to meet at the library to study for the math final. You are 45 minutes late. You did not see your classmate's text messages because you were talking on the phone. You were also late because you were on the phone, talking to a friend. You arrive 45 minutes late and you see your classmate waiting for you. What do you say to your classmate? (You start the conversation)

Pretest 3 (Asking for a Favor):

ASU Student You are in class and a classmate needs to catch up on information they missed the last class. This person wants to see your notes. They are asking you for a favor. You can agree to share your notes or not, depending on how the GL student asks. (GL student starts the conversation)

Global Launch Student: You missed the last class. It was a very important class. The ASU student has good notes. You want to borrow their notes. What do you say to the ASU student? (You start the conversation)

Pretest 4 (Asking about personal life):

ASU Student: You are with a group of friends. You know two people in the group, but the other person (GL student) is new to you. The GL student wants to get to know you. You mention to the group that you have to leave early to pick up your kids from school. The GL student is curious about your children. (You start the conversation)

Global Launch Student: You are with a group of friends. There is one person in the group who you do not know. You are curious about this person (ASU student). The ASU student says they must leave early to pick up their kids from school. You also have

children and you are curious about this person's children. It looks like you may have something in common. What do you say to the ASU student? (ASU student starts the conversation).

Role-Play Posttests

Posttest 1: Declining an Invitation

ASU Student: You invite your classmate to go to a Diamondbacks baseball game with you and your friends next Saturday. You love baseball and you assume that everyone else loves baseball too. You want to invite your classmate because you have an extra ticket and you know that this person has never been to a baseball game before. (You start the conversation)

Global Launch Student: Your classmate invites you to a baseball game with a group of friends next Saturday. You don't understand the sport of baseball and you are not interested. You need to decline the invitation politely. What do you say to your classmate? (Person A speaks first)

roommates? (You start the conversation).

Posttest 2: Apologizing for being late

ASU Student: Your friend does not have a car, but you do, so you made plans to pick this person up from their apartment at 8:00 pm on Friday night to go see a movie. You enter the apartment complex, and you park close to the building where your friend lives. You text your friend to let them know you have arrived and they text back saying that they are coming. Five minutes go by and your friend is still inside. You can't go knock on their door because you do not know which apartment number is theirs. You send a text message and they do not respond. Then, 10 more minutes pass. You honk the horn and then they come out. You are almost late for the movie because you had to wait 15 minutes for your friend to come out. (Person B speaks first)

Global Launch Student: You made plans with a friend to go see a movie on Friday night. You do not have a car, so your friend agreed to pick you up. Your friend said they will arrive at 8:00 pm. You took a shower and then noticed that it was already 8:00 and your friend was waiting for you outside. You answered their text saying you are coming. You still need to get dressed and ready. You take 15 more minutes to get ready while your friend is waiting for you outside. You don't see your friend's text messages because you are busy getting ready. You come out at 8:15 and you run to your friend's car. What do you say to your friend? (You start the conversation)

Posttest 3: Asking for a favor

Instructions: Read the following scenarios and complete the suggested task for Person B. Person A is a native speaker and domestic student and Person B is an international student who is learning English.

ASU Student: You are at your apartment and your other roommates and friends are there. One of your roommates is leaving today to go home for one month. This person needs a ride to the airport and help taking their luggage down the stairs. (Person B speaks first)

Global Launch Student: You are at your apartment. Your roommates and friends are there. Today you are going home for one month and you have to catch a plane. You need a ride to the airport and you need help taking your luggage down the stairs. What do you say to your Posttest

Posttest 4: Asking about private life politely

ASU Student: It is the first week of class and your teacher puts you in a group to do group work. Your group finishes the assignment early. You have time to chat with your group members and when you are chatting you mention that your spouse works at the university. Some of your group members seem surprised to hear that you are married. A group member (Global Launch student) is curious about your age and private life, so they ask you some questions. How do you respond to the Global Launch student? (You start the conversation by mentioning that your spouse works at the university)

Global Launch Student: It is the first week of class and your teacher puts you in a group to do group work. Your group finishes the assignment early. You have time to chat with your group members and you want to make friends with them. One group member says that their spouse works at the university. You are surprised that your classmate is married because you think they are young. You are curious to know more about this person's private life and age. You want to be polite and make friends with this person. What do you say to this person? (ASU student speaks first)

APPENDIX C
RATING SCALES

Appropriateness Rating Scale

5	<p>Excellent - Almost perfectly appropriate and effective in the level of directness, politeness, and formality based on the given context. Expressions are consistently target-like based on the given context.</p> <p>Performed all expected moves for the role-play task.</p>
4	<p>Good - not perfect but adequately appropriate in the level of directness, politeness and formality based on the given context. Expressions are not quite target-like, but pretty good. Performed most of the expected moves for the role-play task.</p>
3	<p>Fair - Somewhat appropriate in the level of directness, politeness, and formality based on the given context. Expressions are more direct or indirect than the situation likely requires. Performed some of the commonly expected moves for the role-play task.</p>
2	<p>Poor- Mostly inappropriate. In terms of directness, politeness and formality based on the given context, expressions could possibly be interpreted as rude or too demanding. Performed only one of the commonly expected moves for the role-play task.</p>
1	<p>Very poor - Clearly inappropriate. In terms of directness, politeness and formality based on the given context. Expressions would certainly be interpreted as rude or too demanding. Did not perform any of the commonly expected moves for the role-play task.</p>
0	<p>No response.</p>

Grammaticality Rating Scale

5	Excellent - Almost perfect. There are no grammatical and/or lexical errors. Extended range of grammatical/lexical variation.
4	Good - There are one or two grammatical and/or lexical errors, which are hardly noticeable, but they do not interfere with understanding. Range of grammatical/lexical variation is good.
3	Fair - There are a few grammatical and/or lexical errors (more than two minor errors or 1-2 major errors), but they do not interfere with understanding. Low range of grammatical/lexical variation.
2	Poor- There are many grammatical and/or lexical errors (several minor errors and/or more than two major errors) that are likely to cause difficulty in understanding. Very limited range of grammatical and lexical variation.
1	Very poor - Expression largely incomprehensible due to excessive grammatical and/or lexical errors. No variation of lexical and grammatical variation.
0	No response.

APPENDIX D
SURVEY QUESTIONS

Attitudinal Survey: Using VR in the Classroom

Description: The following is a 9-question survey that asks about your experience using Virtual Reality in the classroom. Please answer the questions to the best of your ability. The survey should take you approximately 5-7 minutes to complete.

Question 1

Instructions: On a scale of 1-4, with 1 being strongly disagree and 4 being strongly agree, select your level of agreement for the following statement:

1. My experience using Virtual Reality in the English classroom was great!

4- strongly agree, 3- agree, 2- disagree, 1- strongly disagree

Question 2

Instructions: On a scale of 1-4, with 1 being strongly disagree and 4 being strongly agree, select your level of agreement for the following statement:

2. The Virtual Reality activities helped me practice speaking English.

4- strongly agree, 3- agree, 2- disagree, 1- strongly disagree

Question 3

Instructions: On a scale of 1-4, with 1 being strongly disagree and 4 being strongly agree, select your level of agreement for the following statement:

3. I feel more confident in my English-speaking skills after completing the Virtual Reality activities.

4- strongly agree, 3- agree, 2- disagree, 1- strongly disagree

Question 4

Instructions: On a scale of 1-4, with 1 being strongly disagree and 4 being strongly agree, select your level of agreement for the following statement:

4. Global Launch should offer Virtual Reality to all students, every session.

4- strongly agree, 3- agree, 2- disagree, 1- strongly disagree

Question 5

Instructions: Select one of the following answers.

Which of the following VR lessons did you like the most?

- a. Asking for a favor
- b. Declining an invitation
- c. Asking about private life politely
- d. Apologizing for being late

Question 6

Instructions: Select one of the following answers.

Did you like the real-life videos in the software we used, or would you prefer a game-like experience with avatars?

- a. I like the real-life software we used in class.
- b. I would prefer an avatar(cartoon) with an animated (game) environment.
- c. I do not have a preference.

Question 7

Instructions: Type your answer to the following question.

What was your favorite thing about learning with Virtual Reality?

Question 8

Instructions: Type your answer to the following question.

What were the problems with using Virtual Reality?

Question 9

Instructions: Type your answer to the following question.

Do you have any recommendations for using Virtual Reality in future courses?

APPENDIX E
SURVEY RESPONSES

Question 7: What was your favorite thing about learning with Virtual Reality?

- (1) I liked the most the privacy while I answered the questions, and I felt very confident, especially because the answer was written in front of me.
- (2) I like to practice conversations with AI the most because it makes me feel awkward talking to a real person.
- (3) I can practice pronouncing word exactly, and I can get my score immediately after speaking. Besides, there are different levels to practice which is dependent on your ability.
- (4) Using the virtual reality machine is excited. Because I've not used that machine.
- (5) I didn't like it, to be honest, it has a lot of problems.
- (6) Listening from people talking
- (7) My favorite thing about learning with Virtual Reality was that I don't have to take care of mistakes. In addition, I was able to practice pronunciation.
- (8) Learning the daily conversations
- (9) It's a memorable experience.
- (10) There are no favorite thing I prefer to talk to other person it is better than VR.
- (11) I may be very interested in communication through virtual learning, because it makes me feel magical and helpful. This is helpful for me to improve my English communication skills. I have always believed that interest is the best

teacher, and of course it is fresh. It's not a bad thing, it will make me study more seriously.

(12) I can practice many times with VR and don't worry about the other person's mood (Because we are taking with VR).

(13) The most impressive thing is that I can use technology to access information. For instance, we can use VR to interact with people who respond to us. Another thing is that I believe we can learn from lessons through fun games. I think that would be a great way to learn!

(14) when I talked to the VR

(15) You feel comfortable speaking and speak slowly and do not get nervous

(16) The favorite thing was that i can talked without feel nervous because i know it is not a real person.

(17) I enjoyed learning English like a game.

(18) I favor that i have a sense of speak with other person in a virtual place. That is amazing for me.

(19) It is very interesting, just like talking with real people.

(20) actually i do not like vr.

(21) Don't worry about saying the wrong thing

(22) A new way to learn that has never been seen before

(23) It was good to be able to listen to the exact pronunciation and copy it. I was able to speak with accurate pronunciation.

(24) I think the favourite thing about learning with VR is simulates reality and to gave all students time to speak and spill some words that is little hard for us.

(25) virtual reality is fun, but do the same thing repeat and repeat is boring.

Themes:

Novelty (Responses: 4, 9, 11, 17, 22, 25). Feedback and practice (Responses: 2, 3, 8, 12, 14). Helps with pronunciation (Responses: 3, 7, 23, 24). Realistic/imitates real life (Responses: 24, 18, 13). Helps with anxiety (Responses: 21, 15, 16). Negative comment/no favorite thing to report (Responses: 20, 5, 10). Privacy (Responses: 1). Note: Response #24 was repeated for two themes).

Question 8 - What were the problems with using Virtual Reality?

(1) Once I felt dizzy after using VR.

(2) Sometimes the earphone recordings don't go in

(3) Sometimes, the recorder cannot work so it takes long time to do it.

(4) Sometimes, the machine cannot record my voice.

(5) While using it the screen turned black, and I had to press the button every time. There are not working.

(6) The problem with Virtual Reality was that I couldn't start soon.

(7) Sometimes I repeat my sentences a lot but it did not work

(8) There is sometimes a lag or the system receives my voice input but doesn't respond.

(9) Take long time to fix it also it turn off suddenly, sometimes I talk many times and it can understand me.

(10) I think it is a bit monotonous in terms of use in our classroom. For example, we will only use it to do the day's topics and talk to virtual characters. I think this

is not fully utilized. We should experience more things through VR, which is virtual. Related topics in the environment and physical interaction with virtual characters are important in my opinion.

(11) Because the questions from VR are more standardized and less humanistic. So I can't leave a deep impression.

(12) I think the connections sometimes don't work properly. For example, when I record my speaking, it doesn't always catch my voice and has a delayed reaction. That feels really frustrating when you have to repeat yourself and struggle to record your voice.

(13) I think we should take class VR and the other class next week's talk with people because when we repeat the same thing every week it is boring.

(14) Annoys people with long hair

(15) Sometimes I did not understand some students because of the accent. Also, one time I want to practice and my VR was have a problem so I did not practice that day.

(16) Sometimes the microphone doesn't record well and it takes time to move on.

(17) I am not very skilled to use the VR device.

(18) VR is a little bit heavy. I need to use hand to hold the VR even if I adjust the VR size.

(19) VR cannot understand my pronunciation at all.

(20) Sometimes VR software has some programming problems

(21) It's difficult to use.

(22) AI does not understand our English.

(23) It is impossible to know which word or part of it is pronounced incorrectly.

(24) The problem is sometimes the VR does not work easily hard to press the buttons until someone helped me, so that's my time lose.

(25) No.

Themes: Issues with audio input and/or output (Responses: 2, 3, 4, 6, 8, 9, 10, 13, 15, 16, 19, 22) Issues with hardware (VR headset) issues (Responses: 5, 9, 14, 18, 21, 24). Issues with the design of the software (Responses: 11, 20, 23). Issues with implementation (Responses: 10, 13), Dizziness (Response: 1), No issues (Response: 25).

Question 9- Do you have any recommendations for using Virtual Reality in future courses?

(1) It's will help us to pronounce the sentence correctly, so I think would be nice to be a class for speaking that using the VR to work on our pronunciation and speaking skills.

(2) No.

(3) It should have a sample speech so that learners can correct and imitate the true pronunciation.

(4) No, virtual reality lesson was very good.

(5) Maybe talking with real people is better, the doctor and the staff helping her like Jesus and Harry were very kind. I liked talking with them.

(6) Talking fluently

(7) Yes, students should experience both with Virtual Reality and with other students.

- (8) I hope become more flexible to use it
- (9) I would prefer a cartoon with an game environment.
- (10) The recommendation is to not use it in future. Or you must to increase its ability to understand different accents.
- (11) I hope that in future virtual reality classes our class topics can be diversified. The opposite is to add character interaction links and communication that is not subject to restrictions. We can actively communicate with virtual characters. I think this is more important for improving English communication skills. And it will be very helpful.
- (12) Maybe we can do some group games with VR?
- (13) From my perspective, I believe that including video games or teamwork activities in Virtual Reality in future courses would be beneficial and supportive.
- (14) I think it is a great idea to use VR
- (15) No, you are doing your duty perfectly
- (16) Yes i have, may be you can do it more fun. For example, use a video games so that means we can play and learn at the same time.
- (17) nothing in particular.
- (18) I can learn more things about VR device basically.
- (19) I think it is great.
- (20) no
- (21) Hopefully, there will be fewer procedural problems and the dialogue will be more intelligent

(22) More difficult if not easier to handle.

(23) Studying English is fun, but I thought it would be nice to experience VR games

(24) My recommendations for using Virtual Reality in future courses is to verify to another kind that man or to meet someone on street or restaurant, that is more reality.

(25) VR sometimes cannot distinguish other answers, causing us to repeatedly repeat the same content.

Themes: Improve technical issues (Responses: 8, 10, 21, 25) , Content (Responses: 24, 22, 11) , Gamification (Responses: 23, 16, 13, 12, 9), Use for speaking/pronunciation (Responses: 1, 3, 6), Do not use it again (Responses: 5, 10). No recommendations/continue use (Responses: 20, 19, 17, 15, 14, 7, 4, 2)

APPENDIX F
PRACTICE ROLE-PLAY SCRIPTS

Task 1: Declining an Invitation

Hi, how's it going?

- Pretty good, and you?

Not bad. Do you want to come shopping with me on Saturday?

- Come on, you know I hate shopping!
- I'd love to, but I don't have any money at the moment.
- Sorry, I went shopping just yesterday. Maybe we can do something else.

Ok, shall we go to the beach then?

- Yes, that sounds good. Which day?

How about Sunday afternoon?

- I can't, I have tutoring classes.
- I can't, it's my cousin's birthday.
- I can't. I have a family lunch on Sunday.

Well, Saturday then?

- Saturday's good. What time?

We can meet at 11 am.

- Oh sorry. I have a dance class.
- Oh sorry, I just remembered, I have to visit my grandma.
- Oh sorry, I've already got plans for the morning.

Haha, this is so complicated! Maybe we should leave it until next week.

- Yeah sorry, maybe next weekend.

Task 2: Apologizing for Being Late

Hi, how are you?

- Hi, good. Sorry I'm late.

It's all good.

- There was a lot of traffic.
- I overslept.
- The metro was delayed.

Don't worry, these things happen.

- I hope you haven't been waiting for too long.

Not at all, I've just arrived.

- Oh that's good, I was getting worried.

You're fine. Tracy's also late, she just texted me.

- Oh, did she say why?

No, I guess we'll find out when we see her. Maybe we can grab a drink first.

- Sounds good. I'll get it, what would you like?

Oh thanks. Just a coke for me, please.

- Sure, be right back.

Task 3: Asking for a Favor

Hi, are you ok there?

- Hi, could you give me a hand? I can't get this door to open.
- Hi, would you mind giving me a hand with this box? It's a bit heavy.
- Hi, could you hold this for me for a moment? I just need to get my keys out of my bag.

Sure. There we go.

- Thank you so much. Can I possibly ask you for another favor?

Yes, sure.

- Do you happen to have any change with you? I need some coins to pay for the parking, but I've only got these 10 dollars.

Let me check... yes, here you go!

Task 4: Asking about Private Life Politely

Hello Mark, how's your day going?

- Good, thanks. And you?

Pretty good. I have to leave early this afternoon.

- Oh, how come?
- Oh, where are you going?
- Oh, why do you need to leave early?

I just need to pick my son up from school at 3:30.

- I didn't know you had a son. How old is he?
- Oh, you have a son? How old is he?
- Oh, how old is your little boy?

He's eight.

- How cute. Is he your only child?
- How cute. Do you have any other children?
- How cute. Does he have any brothers or sisters?

Yes, I also have a little girl. She's only 3.

- Oh, how sweet.
- Oh, how lovely.
- Oh, it must be nice to have two kids.

Yes, they keep me busy! Anyway, I have to go. Have a good night!

- Thanks, you too.