Foreign Language Acquisition via Artificial Intelligence and Extended Reality: Design and Evaluation

Rahul R. Divekar[1], Jaimie Drozdal[1], Samuel Chabot[1], Yalun Zhou, Hui Su, Yue Chen, Houming Zhu, James A. Hendler, Jonas Braasch

Rensselaer Polytechnic Institute, Troy NY. USA

[1]Equal Authors

The Version of Record of this manuscript has been published and is available in *Computer Assisted Language Learning (CALL)* 15 Feb 2021. https://www.tandfonline.com/doi/full/10.1080/09588221.2021.1879162

Abstract

Artificial Intelligence (AI) and Extended Reality (XR) have been employed in several foreign language education applications to increase the availability of experiential learning methods akin to international immersion programs. However, research in multi-modal spoken dialogue in L2 combined with immersive technologies and collaborative learning is thin, limiting students' experiences to solo interactions focused mostly on vocabulary and grammar in such settings. We intend to fill this gap as we present the Cognitive Immersive Language Learning Environment (CILLE). The AI in CILLE can hear, see, and understand its users and can engage with them in non-dyadic multimodal conversations. The XR offers students a feeling of being somewhere else without the use of intrusive devices and supports multi-party, multi-modal interactions. Together, AI and XR create naturalistic conversational interactions targeted towards comprehensive foreign language acquisition. We evaluate CILLE as a Chinese-as-a-foreign-language (CFL) education tool through a seven-week, mixed-methods study with university students (N=10). Results display statistical significance and retained improvement in CFL vocabulary, comprehension, and conversation skills. Coupled with an analysis of student feedback and researcher observations, we show how CILLE is designed and experienced by students to learn CFL.

Keywords—Foreign Language Acquisition; Augmented, Virtual and Extended Reality; Artificial Intelligence; Spoken Dialogue Systems; Multi-modal Interaction; Multi-party Interaction; Human-computer Interaction; Chatbots; Conversational Agents

1 Introduction

Conversational role-play is a common foreign language learning approach but has two major shortcomings. One, the physical environment rarely matches the conversational context. For example, learners may role-play buying fruits in the target language (TL) in the comfort of their white-walled classroom without the context of an overwhelming, crowded bazaar. Two, conversational partners beyond peers are difficult to find, thereby limiting how much a student can practice. Foreign language learners find themselves asking, "Where and with whom can I try out my new language in the real-world"? Foreign language immersion programs not only provide opportunities for students to travel to a target country for exposure to language and culture, but also make learning less accessible.

We present interactive conversational AI agents that can uniquely expose students to authentic spoken conversations in an XR environment and provide the strengths of international immersion programs without requiring travel (Freed, Segalowitz, and Dewey 2004; Reuland et al. 2012; Taguchi 2015).

Our project is situated in the computer-assisted language learning (CALL) domain which attempts to provide foreign language (FL) exposure, practice, and experiential learning opportunities through technology that span software and hardware solutions (Hubbard 2006), particularly using interactive conversational agents (Hassani, Nahvi, and Ahmadi 2016). Combined with virtual immersion via extended reality, these agents have the power to create authentic spoken interactions for students. From here, we discuss the following themes of literature: Interactive Conversational Agents enabled by Artificial Intelligence and Extended Reality.

1.1 Conversational AI Agents in CALL

Interactive conversational AI Agents (or chatbots) have come a long way in FL education (Fryer, Coniam, Carpenter and Lăpușneanu 2020) since theorized (Fryer and Carpenter 2006), but have been deemed insufficient (Coniam 2008) because of their rigidity and lack of response, among other things. (Bibauw, François, and Desmet 2019) provide a meta-analysis and classification of their purposes in L2 education.

Beyond being an available resource for conversational practice, positive effects of agents have been noted on vocabulary acquisition (Legault et al. 2019), cultural learning (Cheng, Yang, and Andersen 2017), and improved willingness to communicate (Ayedoun, Hayashi, and Seta 2015). They allow independence of use, and exposure to new forms of language (Gallacher, Thompson, and Howarth 2018). They provide a student with authentic material for study and opportunity for experiential learning and improve motivation in students (Hassani, Nahvi, and Ahmadi 2016). They additionally offer the future potential for tailored relationships with students (Edwards et al. 2018). Overall, they are a useful language learning tool (Morton, Gunson, and Jack 2012; Morton and Mervyn 2007). Of the many agents that exist, the most immersive and realistic role-playing experiences enable speech with Automatic Speech Recognition (ASR) and allow the learner to engage in a spoken-dialogue role-play pertaining to a given scenario (van Doremalen et al. 2016; Morton, Gunson, and Jack 2012; Anderson et al. 2008); that is our focus.

1.2 Extended Reality in CALL

Role-play interactions mediated by technology often make use of 3D interactive virtual environments (VE) to better immerse the learner. They are presented in virtual reality (VR), "augmented reality (AR), mixed reality (MR), virtual habitats presented via desktops, and some 3D videos and video games" (Legault et al. 2019). Extended reality (XR) is a term encompassing the newest media for VEs that aim to create a sense of presence and immerse users in a VE.

XR can create opportunities for situated learning and provide a more meaningful connection with learning material and the community (Lave and Wenger 1991; Brown, Collins, and Duguid 1988). XR can do so because it provides an audio-visual dimension with richer contextual and linguistic information than a standard textbook (Yang and Liao 2014). Many studies and meta-reviews note its advantages: increased engagement, authenticity, and confidence (ChengChiang Chen and Kent 2020); improved communicative competence and incidental learning (Yamazaki 2018); greater improvements in speaking performance (Canto and Ondarra 2017; Güzel and Aydin 2016); improved pragmatic understanding (Wang, Grant, and Grist 2020). In addition, Lai and Li (2011) have shown how technology helps Task-based language teaching. Overall, XR provides a superior benefit compared to traditional instruction by a large effect (Tsai and Tsai 2018) with a significant effect on achievements (Avcı, Coklar, and İstanbullu 2019; Chien-pang Wang, Yu-Ju Lan, et al. 2019) and comes with a cost (O'Brien, Levy and Orich 2009).

However, even 30 years after XR's early promises in education articulated by (Bricken 1990), Radianti et al. (2020) note a majority of studies examined make no mention of applied learning theory and focus mostly on evaluating usability or user experience with only 10% evaluating learners' performance through an exam or expert judgment.

1.3 Motivation and Purpose of Study

Recent meta-reviews (e.g, Shawar 2017; Kim, Cha, and Kim 2019; Bibauw, François, and Desmet 2019) and the previous synthesis of the literature suggest that from a language performance perspective most CALL research is geared towards vocabulary or sentence structure education and not conversations. The few that attempt to provide realistic conversational opportunities lack in the following four ways.

(1) AI Agents and XR have rarely been combined to provide authentic conversational exposure while visually immersing users in the relevant context.

(2) Most of the technologies discussed so far assume dyadic (one-to-one) interactions. However, this does not reflect all real-world conversations and current systems and deprive students of exposure to the dynamics of non-dyadic conversations. Enabling a non-dyadic natural conversation with AI Agents is a massive research undertaking as it requires AI Agents to interpret multimodal cues and decide who is speaking to whom and when each agent can interject—all without requiring intrusive wearables.

(3) Technologies that aim to immerse learners in virtual realities typically use a head-mounted device for a greater sense of presence/immersion. They have two key shortcomings: (a) they can cause fatigue/dizziness on prolonged (15-20 mins.) usage (Cheng, Yang, and Andersen 2017), and (b) they can completely remove learners from their physical realities leading to less collaboration with peers in the room. They are significant drawbacks because prolonged practice and collaboration are essential to learning any skill (Sharda et al. 2004).

(4) While we have found, in existing literature, no human-scale collaborative XR CALL environments allowing language learners to interact with multiple simultaneous AI Agents by speaking and hearing responses at the discourse level, the few XR or AI applications in CALL stop before evaluating the educational benefits of applications beyond basic grammar or vocabulary.

Our system utilizes a human-scale panoramic screen to circumscribe groups of Chinese language learners in a virtual environment for a high degree of physical immersion. We combine this with AI Agents that inhabit the panoramic screen and offer conversational roleplay in non-dyadic scenarios. Students interact with them using multiple modalities such as voice, gestures, and transcripts generated by ASR. Such conversations with AI are in line with the *input->process->output* model of learning: students hear passages that reflect real-world contexts and reply to the AI Agents based on what they hear (Ahmadi and Panahandeh 2016). The markerless (i.e. no wearable devices) nature of the system supports multiple simultaneous users who interact naturally with the environment and each other, facilitating the collaborative learning experience. Using our environment, we examine interactive speaking of Mandarin Chinese, rather than solely focusing on individual aspects such as vocabulary.

This paper describes the design and pilot study of our learning paradigm in the AI+XR environment called Cognitive Immersive Language Learning Environment (CILLE). The pilot course delivers content equivalent to one lesson of a standard Chinese Level-2 textbook (Lin 2013). Our study shows a statistically significant effect of CILLE on students' performance through a pre-, post-, and delayed-post- tests design across three broad areas: vocabulary, listening, and speaking (i.e. overall conversation skill). It finds CILLE as a fun, engaging, and comfortable learning space.

2 The Cognitive Immersive Language Learning Environment (CILLE)

CILLIE's hardware consists of a nearly 360° cylindrical panoramic screen measuring 3.8m tall and 12m in diameter, shown in figures 1 and 2. This approach addresses the challenges of traditional VR—front-projection screens cause less fatigue and nausea, and device-less interactions are less intrusive than head-mounted displays (Cruz-Neira et al. 1992). As a result, students can move and collaborate more freely for longer periods of exposure to the learning content; possibly increasing interactions, retention, self-esteem, and above all, preparation for the real-world.

CILLE can "see", "hear", and "speak to" students (i.e recognize multimodal input and produce output) using various sensors and technologies. Kinects placed below the base of the screen recognize gestures and head orientation (R. R. Divekar, Peveler, et al. 2018; Zhao et al. 2018); wireless lapel microphones capture speech; and a multi-channel loudspeaker array emits spatialized audio for congruent audio/visual presentation (that is, the sound is perceived as coming from its corresponding on-screen visual). The audio system also contributes to the situational context of scenes with appropriate ambient soundscapes (e.g. sounds of a crowded street bazaar). The screen itself is made of micro-perforated PVC for acoustic transparency, allowing the loudspeakers to be placed out-of-sight. Interactions enabled as a result can create a sense of presence and allow users to suspend their disbelief.

Students in the environment are immersed in two complementary experiences: the Panoramic Scenes and the Virtual World as shown in Figures 1 and 2, and explained below.

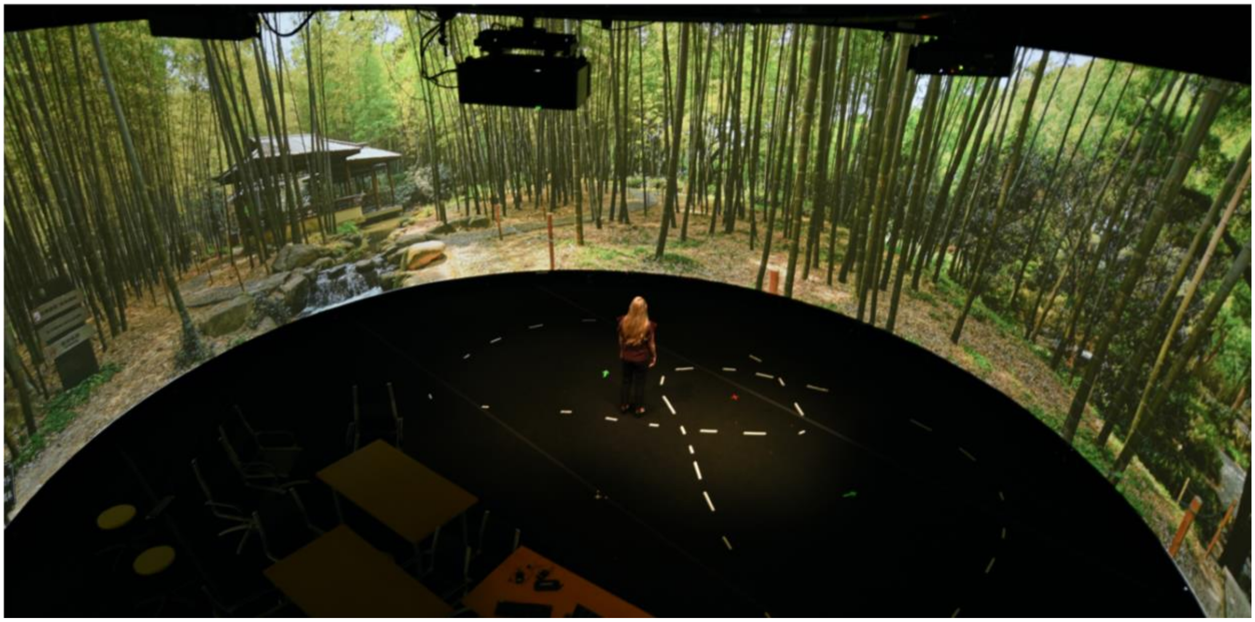


Figure 1: The immersive environment at a university, viewed from above, in which a student experiences a Panoramic Scene of a bamboo forest in China. Best viewed in color

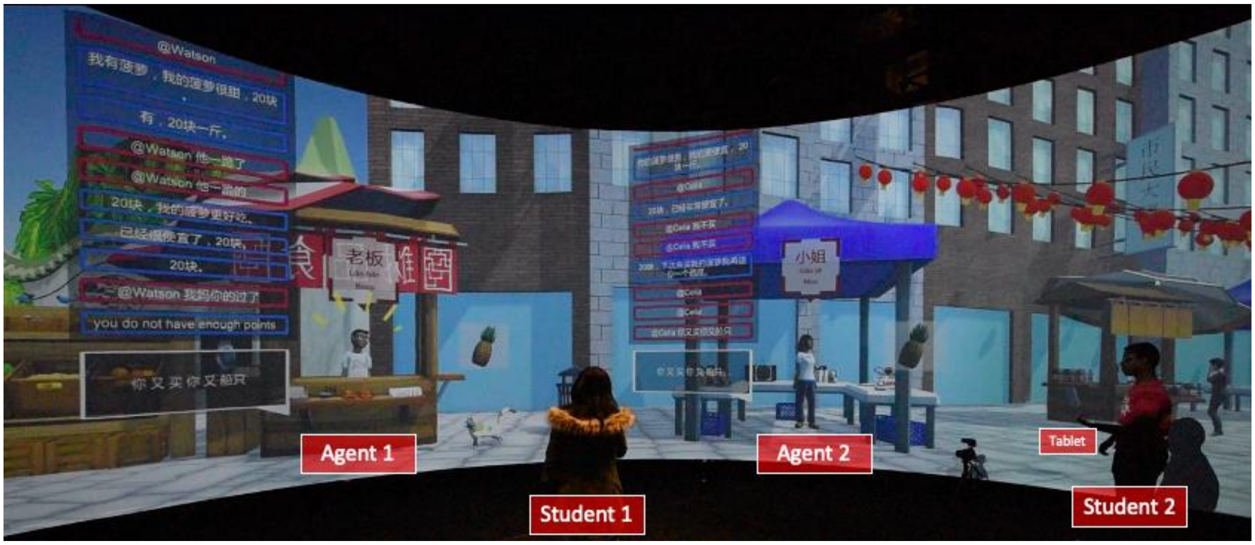
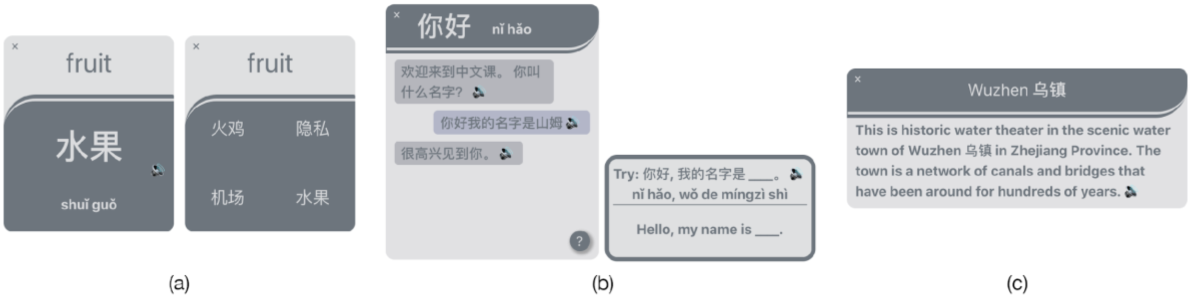


Figure 2: Two students interacting with two embodied AI Agents in the Virtual World scene of a Chinese street market. Best viewed in color

2.1 Panoramic Scenes

Created using real-world 360° images, the Panoramic Scenes focus on acquainting students with vocabulary and cultural knowledge, see Figure 1. Immersion in true-to-life locations enables students to learn and practice amidst the same surroundings they would use such knowledge while traveling abroad. Various interactions facilitate the experience; detailed below and shown in Figure 3.



欢迎来到中文课。你叫什么名字？

你好我的名字是山姆

很高兴见到你。

Try: 你好, 我的名字是 ___。
nǐ hǎo, wǒ de míngzì shì

Hello, my name is ___.

This is historic water theater in the scenic water town of Wuzhen 乌镇 in Zhejiang Province. The town is a network of canals and bridges that have been around for hundreds of years.

Figure 3: Three main types of interactions available in the Panoramic Scenes environment: (a) flashcards and quizzes for vocabulary learning, (b) dialogue boxes (shown with hint tray open) with running transcript between student and AI Agent for speech practice, (c) information cards for conveying language elements or cultural points of interest

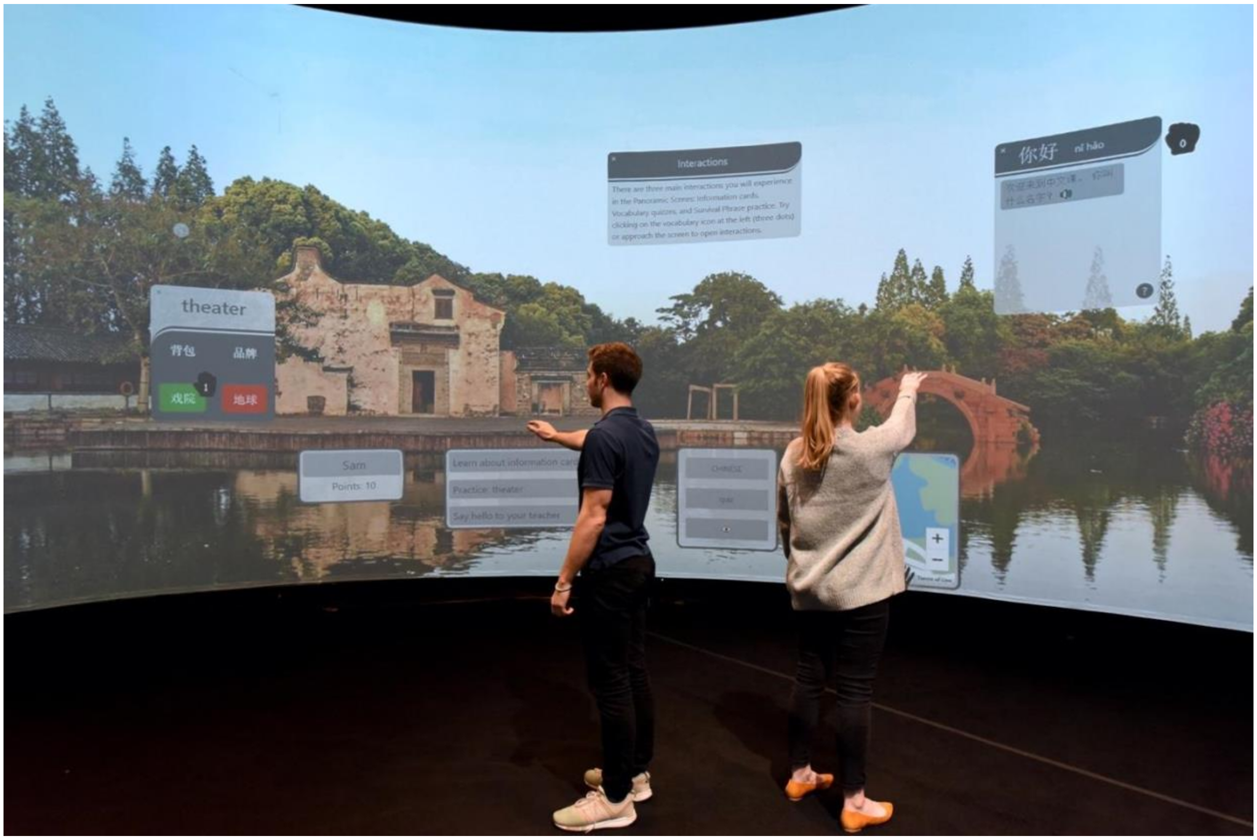


Figure 4: Two users engage with the Panoramic Scene simultaneously for a collaborative exploration of the learning material. Best viewed in color

2.1.1 Cultural Knowledge

An immediate advantage of using rich real-world imagery is the ability to immerse students in authentic cultural experiences, from the familiar to the foreign. Students, in the Virtual World, may visit famous historic sites or witness unfamiliar practices and traditions to glean a better cultural understanding of these locations and improve intercultural competency (Sercu 2010). Interactive information cards embedded (see Figure 3c) in a scene bring cultural knowledge to attention through relevant text about what users may be experiencing. Gesture controls (Zhao et al. 2018; R. R. Divekar, Peveler, et al. 2018) capture hand motions: pointing with a hand and closing into a fist clicks on items in the scene allowing interactions without intrusive devices used commonly in motion tracking. When information cards are clicked on, target words and phrases are read aloud for pronunciation understanding.

2.1.2 Vocabulary Learning

Students can learn and practice target words through an exploration of actual items and objects within the scene in the word's context. These interactive items engage the student in two modes that can be toggled between Explore and Practice. In Explore mode, the student opens flashcards featuring the target word's translation in Hanzi[1] and pronunciation in Pinyin[2]. Practice mode transforms this flashcard into a multiple-

[1] Hanzi are the logograms and most popular way of writing and reading Simplified Chinese

[2] Pinyin is the Romanized spelling of Chinese words that conveys pronunciation and tone

choice quiz with four candidate translations (see Figure 3a). Selecting an answer reads it aloud and reveals its English translation.

2.1.3 Sentence-level Practice

Students engage with *dialogue cards* to practice listening comprehension and speaking. Opening a card reveals a dialog box that prompts the student aloud with a statement or question related to content in the scene. Students can speak their response; what is heard by the system shows on the screen as a text transcript so students can see where they may have difficulty with pronunciation Those struggling to answer the prompt can access assistance by toggling a tray of candidate responses. The system can play the audio for these aloud so users may "listen and repeat" (see Figure 3b). Example dialogue of a card beside a cashier in a scene might go as follows:

Card: 现金还是信用卡？(Cash or Credit?)

Student: 现金 (Cash)

Card: 好的! (Okay!)

2.1.4 Gamification Elements

Incorporating typical game elements, or "gamification,'' is widely employed in CALL as a technique for increasing user motivation and engagement (Hung et al. 2018), two factors known to be important in the language learning process (Flores 2015). Most recently, (Tang and Taguchi 2020) have found success using games to acquaint students with complex knowledge such as culture, pragmatics, etc.

In our Panoramic Scenes, multiple scenes are connected to form "quests'' focusing on a theme i.e. street markets. Quests involve games such as "I Spy'' which prompts the student to identify and select an item in the scene to receive points. The use of AI[3] to create games by automatically classifying images discovered in the scene reduces effort required to create content.

2.2 Virtual World

While the Panoramic Scenes use real-life imagery, the Virtual Word immerses users in computer-generated scenes realized through the Unity game engine, which gives more programmatic control over the visuals. AI Agents here are embodied as humanoid avatars who pose as shopkeepers in a Shanghai street market (shown as Agent 1 and Agent 2 in Figure 2). The users are engaged in extended rounds of conversations with each other and both avatars at the same time.

Users can interact with an agent simply by looking at one and speaking (R. R. Divekar, Kephart, et al. 2019) rather than having to use a typical wake-up word which has been a known detriment in multi-party conversations (R. R. Divekar, Drozdal, et al. 2018). The agents are designed to inflate prices and persuade users to buy their products. Agents can also interject a conversation to compete with each other and get a user's attention (R. R. Divekar, Mou, et al. 2019; Bayser, Pinhanez, et al. 2018; Bayser, Guerra, et al. 2018). In doing so, the two AI Agents can critique the quality of each other's products in comparison to their own, thus exemplifying nuances of a street market. Such a haggling and competitive dynamic is uncommon in a user's home country; the agents are designed to raise intercultural competence through interactive roleplay. The resultant dialogue includes elements of common shopping scenarios in China, both linguistically and culturally. It was created by a diverse group of language and culture experts who are careful to avoid misrepresenting the target culture.

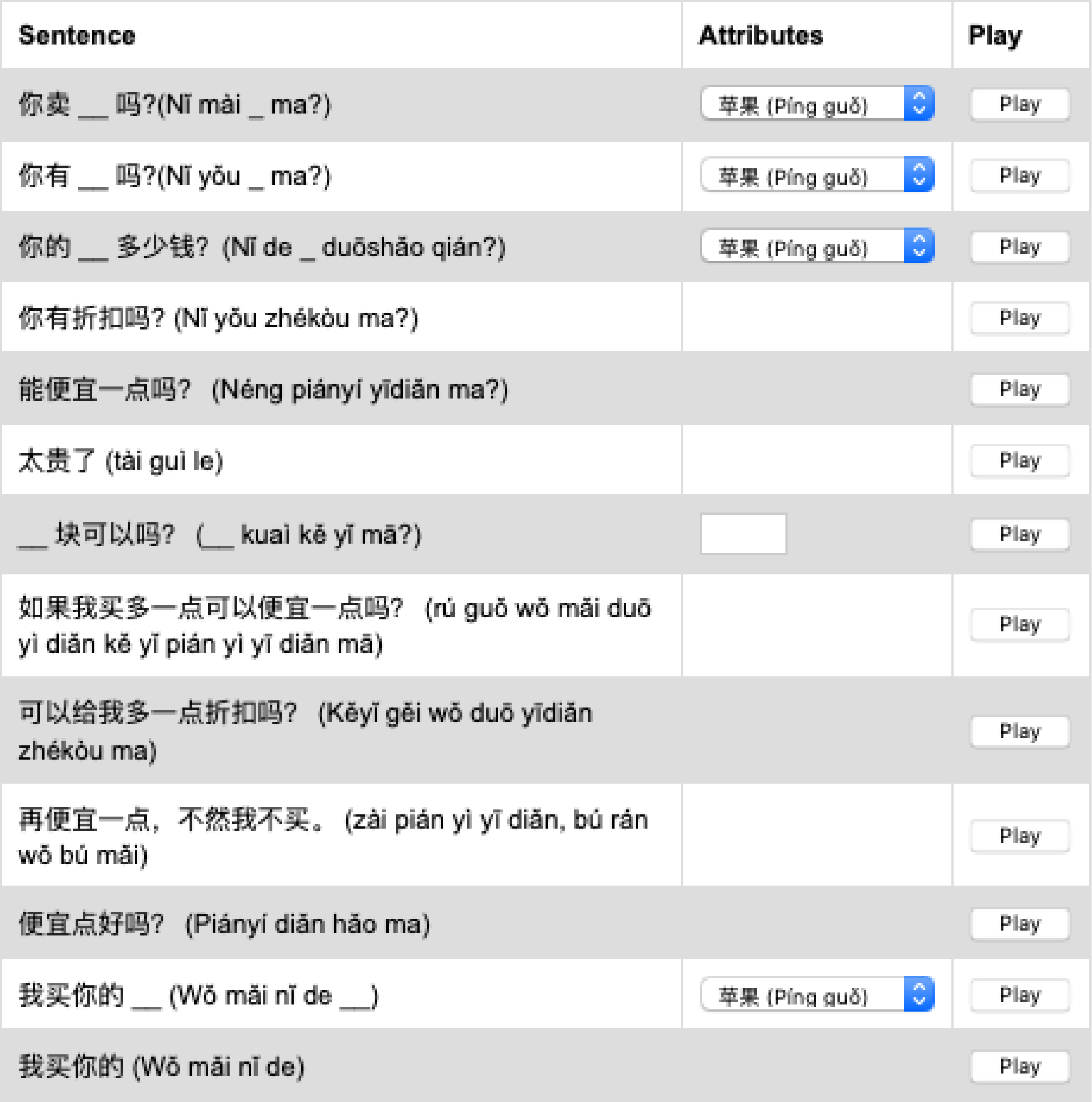
Other features of the interaction include a transcript of all utterances displayed next to the avatars (seen Figure 2) shown with the hope of raising users' *noticing and awareness* as it lets them see in real-time which

[3] https://www.ibm.com/cloud/watson-visual-recognition

characters weren’t correctly pronounced. It also helps users understand what the agent said in case users didn’t fully understand the audio (R. R. Divekar, Zhou, et al. 2018). We included a listen-and-repeat feature for beginner speakers as they need more assistance in roleplay. We implemented it on a tablet as it is unlikely to cause an intrusion to realism given its ubiquity. A screenshot of the tablet UI is shown in Figure 5. Students can select phrases and words they want to listen to, hit play, listen, and repeat the sentence into the mic. Live transcription allows them to notice gaps and continuously improve until they are ready to speak to the agents again.

Overall, the experience is intended for students to learn communication through Task-Based-Language-Teaching (TBLT) where an example task may be to purchase items for a fair price within a budget. Interactions that support this goal were discussed in this section. A simplified demo in English can be found online[4].

[4] Demo Video: https://youtu.be/KBV9z9fLAD0



Sentence	Attributes	Play
你卖 __ 吗?(Nǐ mài _ ma?)	苹果 (Píng guǒ)	Play
你有 __ 吗?(Nǐ yǒu _ ma?)	苹果 (Píng guǒ)	Play
你的 __ 多少钱? (Nǐ de _ duōshǎo qián?)	苹果 (Píng guǒ)	Play
你有折扣吗? (Nǐ yǒu zhékòu ma?)		Play
能便宜一点吗? (Néng piányí yīdiǎn ma?)		Play
太贵了 (tài guì le)		Play
__ 块可以吗? (__ kuai kě yǐ mā?)		Play
如果我买多一点可以便宜一点吗? (rú guǒ wǒ mǎi duō yì diǎn kě yǐ pián yì yī diǎn mā)		Play
可以给我多一点折扣吗? (Kěyǐ gěi wǒ duō yīdiǎn zhékòu ma)		Play
再便宜一点，不然我不买。 (zài pián yì yī diǎn, bú rán wǒ bú mǎi)		Play
便宜点好吗? (Piányí diǎn hǎo ma)		Play
我买你的 __ (Wǒ mǎi nǐ de __)	苹果 (Píng guǒ)	Play
我买你的 (Wǒ mǎi nǐ de)		Play

Figure 5: Screenshot of the Tablet UI. Students can play one of the available phrases (displayed in random order to avoid memorization) to ``listen and learn'' during the interaction.

3 Methodology

Our experiment evaluates the effectiveness of the CILLE through a mixed-methods user study with 10 participants recruited from a local university[5]. All participants were novice-low students. They were

[5] The user study was approved by the university Institutional Review Board (IRB). Participants were compensated for their time with a gift card.

arbitrarily paired, and each pair experienced four 1-hour learning sessions in the immersive environment. The learning sessions were designed to teach Chinese as a Foreign Language (CFL)—specifically, a unit on shopping in street markets in China.

We included a handful of these learning outcomes and evaluation methods acceptable in the relevant literature (Hubbard 2006) in our study design. We measured the effectiveness quantitatively through listening, comprehension, translation, and conversational proficiency scores tracked across each participant's pre-, post-, and delayed post-tests, as well as through students' self-reported learning gains. Qualitative measurements are provided by data gathered from a questionnaire distributed after the study. Additionally, the participants were asked to log entries into a digital diary/journal after each learning session and after exposure to Chinese outside the study sessions to track external effects on scores. The schedule of learning and testing sessions are shown in Table 1.

Table1: Schedule of various stages of the study

Week | Day	Task	Goal
1 | 1	Pre-test	Measure students' CFL knowledge and proficiency before learning sessions.
2 | 1	Learning Session 1	Demonstrate technology. Students engage with and practice target vocabulary and sentence structures in the Panoramic Scenes (30 minutes).
2 | 2	Learning Session 2	Students engage with and practice target vocabulary and sentence structures in the Panoramic Scenes (30 minutes).
3 | 1	Learning Session 3	Demonstration of technology. Students role-play conversational tasks with AI Agents in the Virtual World (40 minutes).
3 | 2	Learning Session 4	Students role-play conversational tasks with AI Agents in the Virtual World (40 minutes).
4 | 1	Post Test	Measure students' CFL knowledge and proficiency after learning sessions.
5	Break	
6	Spring Break	
7 | 1	Delayed Post Test and Feedback+	Measure students' CFL knowledge and proficiency after a three week delay and close the study with feedback from students.

+Conducted via teleconferencing due to Covid-19 pandemic

3.1 Demographic

Participants were university students (N = 10), ages ranged from 18 - 22 years old (*M* = 19.1, *SD* = 1.3). 50% identified as female and 50% identified as male. We established participant eligibility through the Language History Questionnaire (Li, Sepanski, and Zhao 2006), a self-reported measure of previous foreign language experience. Subjects recruited had not been exposed to Chinese in any significant way. Nine participants indicated English as their first language and one indicated Tagalog. Concurrent to the study, participants were enrolled in courses in Level-1 Chinese as a Foreign Language (CFL) at the university. Outside these courses, students indicated minimal exposure to Chinese in day-to-day activities. They noted some usage of popular CALL tools such as Duolingo and Quizlet.

3.2 Learning Outcomes

The learning content for the study included 32 vocabulary words and 6 sentence structures related to shopping for fruits at markets in China. This content, distinct from that in the CFL course students were taking concurrently, is comparable to Unit 6, Lesson 1 of (Lin 2013), a university-standard textbook. The CFL professor consulted throughout the study noted this content typically requires two classroom sessions (approximately 220 minutes) plus homework (240-420 minutes, dependent on student aptitude). In comparison, our user study included four learning sessions and required no homework (a total of 240 minutes).

In addition to language elements such as vocabulary, research (Adair-Hauck et al. 2006) advocates that proficiency, or being able to use the language elements in real-world conversations, presents a more holistic goal of learning. Much of the holistic language learning goals can be captured via can-do statements that layout what students can achieve at the end of a session. In 2017, NCSSFL-ACTFL published guidelines to set learning outcomes via can-do statements and classified them into interpretive, interpersonal, and presentational communication goals[6].

Our can-do statements are also informed by recent research adding culture to get a richer understanding of the target language (Oxford and Gkonou 2018) and to avoid the trap of being a "fluent fool" (Bennett 1997; Choudhury 2013).

Considering this, our study utilized the following interpretive and interpersonal can-do statements as learning outcomes. After the study, participants should be able to perform the following: I CAN:

- Interpret/identify the price of an object at a Chinese street market
- Interpret/identify various qualities of objects at a Chinese street market
- Interpret phrases that vendors use to compete and negotiate
- Understand the cultural differences and appropriate ways to bargain in China
- Express wanting to buy something in Chinese
- Ask for availability of items in Chinese
- Inquire about the price of items and negotiate with street vendors in Chinese

3.3 Learning Sessions

Participants were grouped in pairs for learning sessions and each pair came to the study twice a week, one pair at a time for four 1-hour learning sessions. To ensure comparable learning experiences across pairs of participants, sessions were organized into timed segments.

3.3.1 Learning Sessions 1 and 2

Learning sessions 1 and 2 allowed students to engage with and practice the target vocabulary and sentence structures. After an interactive demonstration of the technology, an 8–10 minute slideshow was displayed on the 360° screen to introduce the content for that session. The slideshow included pre-recorded explanations that encouraged students to "listen and repeat"—AI was not yet involved. A corresponding handout was given to the participants to follow along and to take home. After reviewing the new content pairs were given 30 minutes to explore the different interactions in *Panoramic Scenes* (see Section 2.1) that aimed at reinforcing knowledge. At the end of these learning sessions, a short quiz was administered via a laptop. Quizzes were informally used to measure student progress and did not provide any feedback to the learner.

[6] National Council of State Supervisors for Languages (NCSSFL) and the American Council on the Teaching of Foreign Languages (ACTFL) can do statements: https://www.actfl.org/publications/guidelines-and-manuals/ncssfl-actfl-can-do-statements

3.3.2 Learning Sessions 3 and 4

Learning sessions 3 and 4 employed the *Virtual World* (see Section 2.2) and were geared towards using the building blocks of language to complete conversational tasks related to the “can-do statements” (outlined in Section 3.2). Sessions extended the participants’ declarative knowledge into procedural CFL knowledge (Dörnyei 2009).

Each of these sessions began with a self-review of handouts and a 10 minute pre-recorded slideshow review of key sentence structures. Students were given a demonstration before their first interaction. Following Task-based Language Teaching (TBLT) pedagogy, participants were given the goal, or task, to buy fruits in the virtual world with the points collected from learning sessions 1 and 2. One student spoke to the AI Agents while their partner operated the tablet (Figure 5) to provide help. After 20 minutes, the students switched roles and continued the interaction for 20 more minutes. Students were expected to speak to each other, ask for and offer help, discuss what they wanted to say to the AI Agents, and play sample phrases so the speaking student can “listen and repeat.” A sample dialogue can be seen in Table 2.

None of the four learning sessions involved language-expert or human intervention beyond fixing bugs.

Table 2: Sample dialogue between two students and two AI Agents in the virtual Chinese street market. Notice that agents speak only at appropriate turns.

Speaker	Intended Addressee	Utterance	Intended Meaning of Utterance
Student 1	Agent 1	我想卖西瓜	The student wants to say “I want to buy a watermelon” but because the words “buy” and “sell” differ only in tone, the student ends up saying “I want to sell a watermelon”
Student 1	Student 2	Looks like I got that character wrong. Can you play the sound for it?	Student 1 asks for pronunciation help from Student 2 who holds the tablet UI
Student 2	Student 1	Here you go	The student has selected the right sentence and attribute on the UI. The sound of 我想买西瓜 plays on the speakers.
Student 1	Agent 1	我想买西瓜	The Sstudent gets it right^{+}. Asks “I want to buy watermelon”
Agent 1		来看看我的西瓜，很好吃，10 块。	Since Agent 1 was looked at directly, it gets to respond first. It says “Here, look at this watermelon. It’s delicious. 10 kuai!”
Agent 2		他的西瓜不甜，我的很甜，只要8块。	Agent 2 jumps in opportunistically and says “His are not sweet. Mine are. Only 8 kuai!”

+ We saw that students went back and forth between listening, repeating and contrasting what they say with its transcription on screen several times before they got it right; thereby noticing and then correcting the gap.

3.4 Evaluation Methods

In this mixed-methods study, we evaluate our system from two perspectives – How the students fared in the language tests and how they felt about their learning experience. For the former, each participant was administered tests before (pre-), after (post-), and 3 weeks after (delay post-) the last learning session. For the latter, a questionnaire was given to each student.

3.4.1 Testing Sessions

Testing is an evolving research area and many test designs are possible. We use multiple choice quizzes (MCQs) and conversational tests to measure the following CFL metrics: vocabulary, listening, comprehension, and conversational proficiency. Testing across these several metrics as opposed to simply one or two is a key contribution of this work. Questions in the pre- and post-tests were mutually exclusive to avoid effects that may be introduced by testing and retesting on the same content. Questions in the delayed-post-test were drawn from the pre- and post-tests. All tests had the same number of questions in each of the four sections. A sample of questions used can be seen in appendix A. We tested for the following:

- **Vocabulary Recognition:** This section determined which target words students could identify in Hanzi and Pinyin (see Appendix A.1).
- **Listening and Transcription:** This section tested how well one is able to discern various words in an audio clip. Students could play the audio clip at most twice and tested in Hanzi and Pinyin (see Appendix A.2).
- **Listening Comprehension:** This test identified if students could understand what they heard in an audio clip. They were tested at the word and discourse-level listening comprehension. (see Appendix A.3)
- **Interactive Conversation:** A semi-structured three-part conversation between a native speaker (fixed for all) and each participant was used to determine CFL proficiency (see Appendix A.4). This conversation was recorded and received individual proficiency scores from two CFL professors, which were averaged. Proficiency scores were calculated using four-point scales across five metrics: Comprehensibility, Fluency, Pronunciation, Quality of Vocabulary, and Quality of Grammar.

3.4.2 Student Experience Questionnaire

We used a questionnaire at the end of the four learning sessions to measure subjective learning gains and collect perceptions about the technology and learning experience. We elicited the following:

- **Self-reported learning gains:** On a set of true/false questions, students were asked whether they learned new vocabulary, grammar, and culture notes. Participants were asked to check off any can-do learning statements they thought they learned. Such a recap of learning outcomes is common in traditional classrooms, where educators, at the end of a session, reiterate or ask about what was learned that day. Students were also asked to rate on Likert Scales their increase in preparedness and confidence to shop in Chinese markets.
- **Student Experience:** Students were asked to indicate on Likert scales how fun and engaging the experience was. Additional questions were also asked about levels of comfort, anxiety, and engagement with AI Agents. The helpfulness of a key learning mechanism, the "listen and repeat" was also rated on Likert scales.
- **Media Alternatives:** Students were asked about their preferences for several media commonly used to learn languages. Each question asked students to pick the preferred of two presented options based on their own previous experiences.

4 Results

4.1 Test Scores

4.1.1 Total Achievement Score

With all test sections combined, students score out of 75 points. The results (see Figure 6) show that total achievement scores for the 10 participants are significantly higher after four learning sessions with large effect size (M = 58.95, SD = 12.69) than before (M = 32.85, SD = 17.6), $t(9)$ = -9.85, p < .001, d = -1.41. The results also suggest the knowledge was retained after three weeks as delayed post-test scores (M = 60.80 , SD = 12.75) are not significantly different from post-test scores $t(9)$ = -1.55, p = .15, d = .14.

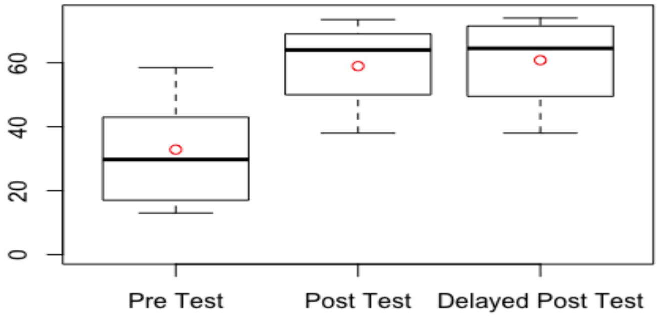
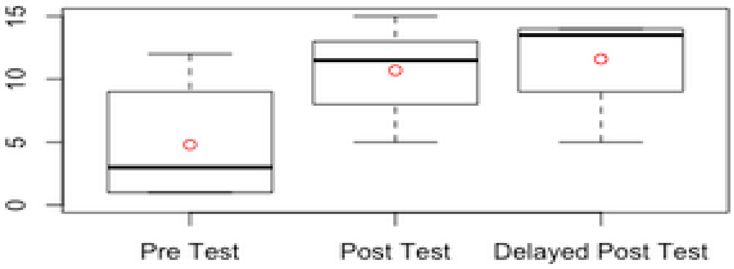


Figure 6: Summary of total achievement scores for 10 CFL learners. The Red dot signifies mean

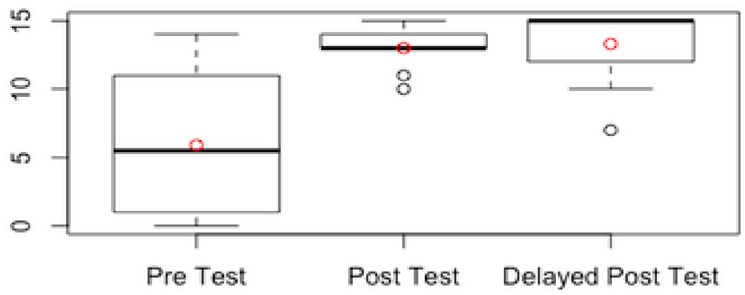
The following sections report statistics on individual sections of the language test using the two-sided Wilcoxon signed-rank test which is more suited for the non-normal distributions that we see in many individual sections.

4.1.2 Vocabulary Recognition

Seen in fig. 7a, Hanzi recognition scores significantly increased from the pre-test (M = 4.8) to the post-test (M = 10.7), V = 55, p = .005. Vocabulary scores for Pinyin recognition (Figure 7b) also increased significantly from the pre-test (M = 5.5) to the post-test (M = 13), V = 55 , p = .005.



(a) Read the English word and pick the correct Hanzi representation (Vocabulary - Hanzi Recognition)

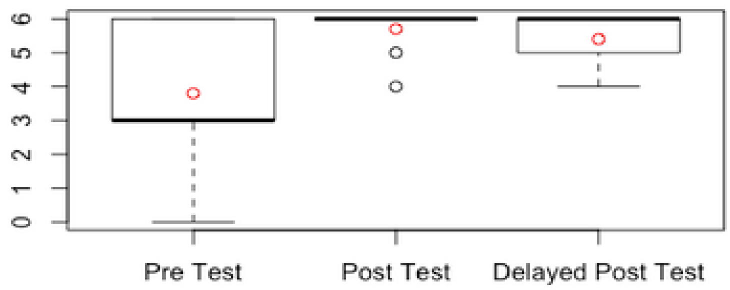


(b) Read the English word and pick the correct Pinyin representation (Vocabulary - Pinyin Recognition)

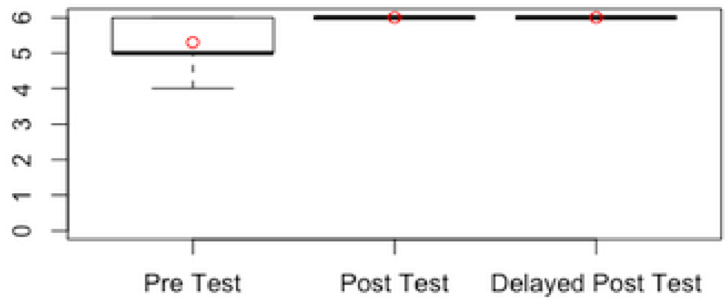
Figure 7: Vocabulary Scores. The Red dot signifies mean

4.1.3 Listening and Transcription

Scores for listening and transcribing to Hanzi (Figure 8a) significantly increased from the pre-test (*M* = 3.8) to the post-test (*M* = 5.7), *V* = 21, p = .03. Similarly, results show a significant increase in scores for listening and transcribing to Pinyin (Figure 8b) from the pre-test (*M* = 5.3) to the post-test (*M* = 6), *V* = 21, *p* = .02.



(a) Pick the correct transcription (in Hanzi) of what you hear

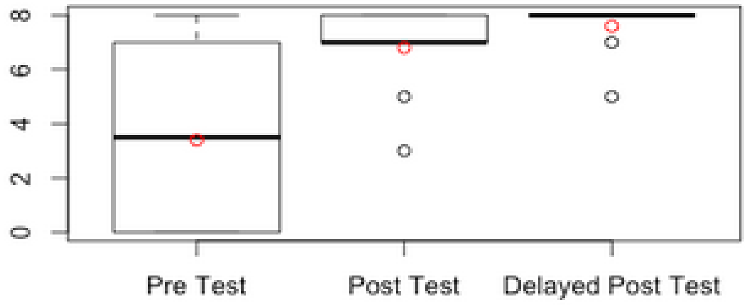


(b) Pick the correct transcription (in Pinyin) of what you hear

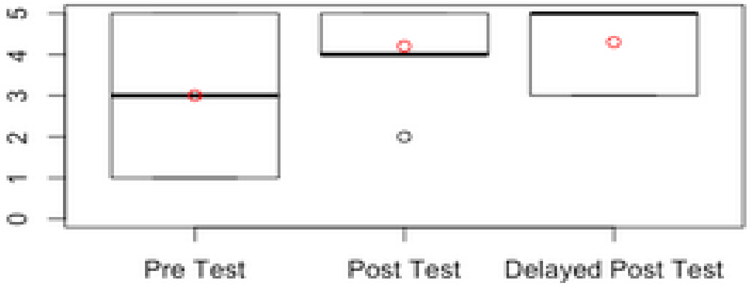
Figure 8: Listening and Transcription Scores. The Red dot signifies mean

4.1.4 Listening Comprehension

Scores for listening and identifying the English translation of a Chinese word (see Figure 9a) varied greatly at the pre-test (*M* = 3.4). However, the range of scores narrow and show a significant improvement of scores after the four learning sessions (*M* = 6.8), *V* = 28, p = .02. Scores for listening to a dialogue and answering related questions (see Figure 9b) are also significantly higher at the post-test (*M* = 4.2) than at the pre-test (*M* = 3.0), *V* = 2.5, p = .03)



(a) Pick the correct translation of the word you hear (Listening Translation - Word Level)



(b) Listening Comprehension - Questions based on the audio clip of a 2-person conversation were asked

Figure 9: Listening Comprehension Scores. The Red dot signifies mean

4.1.5 Conversation

Recorded audio was rated by two experts for CFL proficiency. Inter-rater reliability is high for all three (pre-, post- and delayed-post-) testing sessions (*kappa* = 0.94, *kappa* = 0.97, *kappa* = 0.98, *p* = < .01). The results show a significant improvement in proficiency scores between the pre-test (*M* = 7.05) and the post-test (*M* = 12.55), *V* = 362, *p* = .01. Two participants' conversation test scores did not change from the minimum score between the pre- and post-test, and no participant had a perfect score for all three of the conversation tests. There is no significant difference in proficiency scores between the post- and delayed-post-tests. Because the range of conversation test scores in the post-tests is large (see Figure 10), student responses were isolated and coded into "incorrect," "correct" and "unanswered" through a thematic analysis of the conversations. When the responses from all participants are pooled, in the pre-test only 12% of the responses are correct and 65% of the responses went unanswered. Results from the post-test show a 45% increase in correct responses and a 31% decrease in unanswered responses.

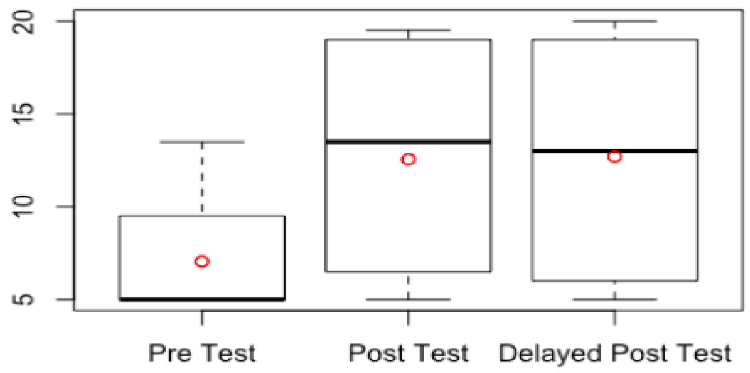


Figure 10: Summary of Conversation Test Scores Across 5 metrics – Comprehensibility, Fluency, Pronunciation, Vocabulary and Grammar (total average of two raters)

4.2 Student Experience Questionnaire

4.2.1 Self-Reported Learning Gains

On a set of true/false questions, 10/10 participants indicated they learned new vocabulary and new sentence structures, and 9/10 indicated they learned new cultural knowledge.

All participants indicated learning at least five of the seven can-do statements; half of the participants indicated all seven (see Figure 11).

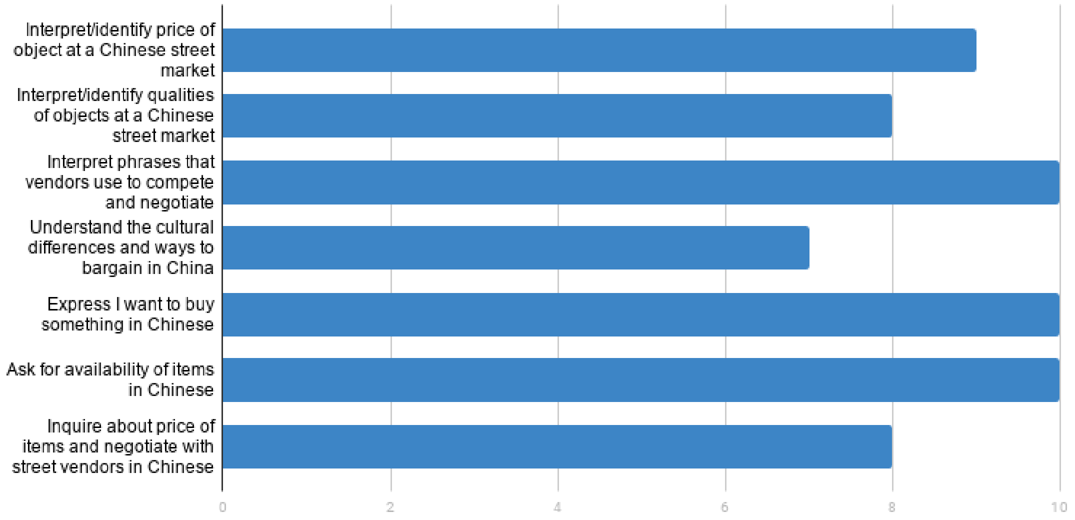


Figure 11: Participants' self-reported responses to each of the seven learning outcomes upon completion of the learning sessions

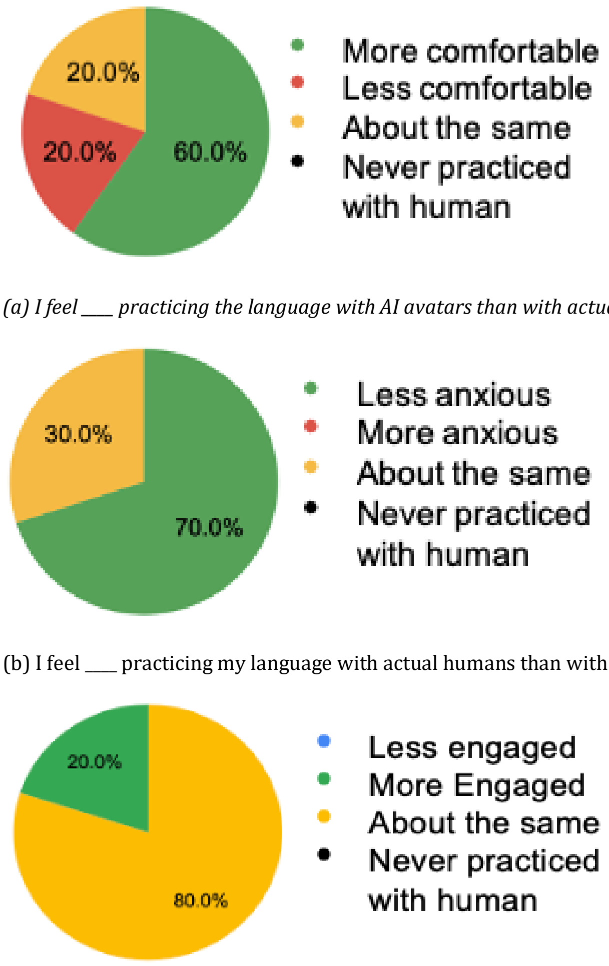
On a 5-point Likert scale of "less prepared" to "more prepared", participants indicated that they feel more prepared to shop (M = 4.5, SD = 0.5) and negotiate (M = 4.3, SD = 0.6) in the street markets of China as compared to when they started.

4.2.2 Student Experience

On two separate 5-point Likert scales, one from "not fun at all" to "very fun" and the other from "not engaging at all" to "very engaging", students expressed that the overall experience was *fun* (*M* = 4.5, *SD* = 0.7) and *engaging* (*M* = 4.6, *SD* = 0.7).

Participants strongly indicated on a 5-point Likert scale that the “listen and repeat” function in Panoramic Scenes (*M* = 4.9, *SD* = 0.3) and the tablet UI in the Virtual World (*M* = 4.2, *SD* = 1) were helpful to their learning success.

Students indicated that they are less anxious, more comfortable, and equally engaged while practicing conversation with AI as compared with humans (see figure 12).



(a) I feel ___ practicing the language with AI avatars than with actual humans

(b) I feel ___ practicing my language with actual humans than with AI avatars

(c) I feel ___ practicing my language with actual humans than with AI avatars

Figure 12: Users’ reflection on interactions with AI Agents. Best view in color

4.2.3 Comparing media alternatives

Questionnaire responses reveal the students’ interests in using the CILLE compared to other media alternatives.

- 80% of participants indicated a preference for the CILLE compared to YouTube videos. (20% equal preference)
- 50% of participants indicated they would use the immersive experience over a traditional classroom. (40% equal preference, 10% traditional classroom)
- 50% of participants indicated that they would elect for the immersive experience over textbooks. (50% equal preference)
- 60% of participants indicated they would use the CILLE over other existing online platforms. (40% equal preference)

5 Discussion

5.1 Learning Gains

The data analysis in Section 4 indicates that students' proficiency across vocabulary, listening, comprehension, and conversation increased significantly between the pre- and post-tests . It shows that their interactions in CILLE were educationally beneficial. There was no significant change between the post- and delay-post-test scores, indicating knowledge and proficiency were retained. Overall, Pinyin was learned more easily than Hanzi. This is expected as Pinyin is Romanized Chinese and reading it is closer to reading English. In conversation tests, along with increased scores, there was a 45% increase in correct responses and a 31% decrease in unanswered questions. For example, in the pre-test when a student was role-playing with the interviewer, most students responded, "I don't know." In the post-test and the delayed-post-test, most students responded in Chinese with contextually correct phrases like "太贵了 (It is too expensive!)" or produce a sentence that was partially accurate. In the self-reflection questionnaire, all students indicated learning new vocabulary and sentence structures. Participants also checked off most of the interpretive and interpersonal can-do outcomes and indicated that they felt more prepared and confident to negotiate in Chinese after their learning experience. This was our goal: to prepare students for conversations in the real-world.

Nine out of ten students indicated learning new cultural knowledge, and many recognized cultural differences and noted these in their questionnaire responses: "In the US, we use coupons instead of asking for discounts," "It is expected to barter [negotiate] in China, while in America it is looked down upon," "Chinese vendors are open to bargaining," "They also have a much higher price markup," etc. One student noted, "items like coke (sic) have a set price while fruit vendors tend to drive up the price and are willing to negotiate."—an important cognizance to not generalize all items as negotiable. While we see anecdotal evidence of cultural learning, a lack of an agreed-upon standard for intercultural competency hindered us from objectively measuring it.

Students expressed that the Panoramic Scenes were helpful in enforcing vocabulary and noted the contextual and cultural relevance of the scenes;the Virtual World was useful for improving conversation and proficiency skills. One student remarked, "...my main takeaway from Panoramic Scenes was the vocabulary in context. The way it had me interacting with the new material built a good foundational association between the words and what they meant." Other students said of the Virtual World "I got to see what it was like to barter [negotiate] for the first time," "I learned... basic conversations that we would have with vendors or store owners when trying to buy goods," and "..it .helped me become more comfortable recognizing phrases and understanding new content in a spoken conversation".

Students, individually and as pairs, leveraged multiple help functions for feedback during their learning. Specifically, researchers observed:

Live Transcript— Students engaged with the transcript and each other to find meaning in the interaction at hand and to clarify actively what they heard. For example, when an AI Agent unexpectedly included a new, free fruit as part of his/her offer, students looked to the transcript for clarification and discussed whether to take the deal. For feedback on pronunciation, students isolated and rehearsed individual words and checked the transcription for confirmation; particularly common for words found to be difficult e.g. 橙子 (oranges) and 西瓜 (watermelon).

Listen and Repeat— The "listen and repeat" help functions available in both learning experiences were also used by all the students and were perceived as greatly helpful to their learning success. In the Virtual World, this help function was integrated into the tablet and required students to collaborate. Students decided together what to say, listened to ideal pronunciations, and spoke to the Agents. By the end of the sessions, pairs combined vocabulary and phrases from their knowledge and the tablet to attempt more complex negotiation phrases, thereby creating meaning together rather than simply repeating material.

5.2 Student Experience

Students were active in their own learning experience and were engaged in the tasks and the immersive surroundings. This was also evident in their interactions, responses to questionnaires, and diary entries throughout the study. For example, researchers observed displays of excitement when a student, often after several tries, got an utterance correct and heard the agent affirmatively respond to them. On the contrary, at times there was visible disappointment in students when they did not get the utterance or interaction right, even after several tries. In these situations, the partner's help was critical.

Researchers observed that some students consistently didn't provide help or collaborate with their partners. This trend was also noted by the students in diary entries. One entry remarks "Not getting much help from [partner] who is mostly watching." The researchers suspect the lower amount of collaboration in some pairs is partially related to the unbalanced CFL knowledge between them at the outset of the learning sessions. In the Panoramic Scenes, it was common for such pairs to explore independently of each other. The high-scoring partner would complete most of the interactions and move ahead to new scenes, thereby limiting the learning opportunities of the struggling student. In the Virtual World, the higher-scoring partner while operating the tablet UI offered less help to the lower-scoring student. Such asymmetric interaction patterns in pairs with different levels of proficiency have been seen previously in a different context of test-taking (Kot Artunç and Ortaçtepe Hart 2020).

Overall, attitudes toward working in pairs of two were mixed: some wanting more, less, or the same number of partners, or a different partner. Perhaps this speaks to the arbitrary pairing process. Three possible, yet difficult, solutions to this could be experimented with in the future. One, students could be paired more methodically. Two, AI could play a more pronounced role in providing feedback. Three, AI may interject to elicit feedback from the students themselves, balancing the reliance between the partner(s) and system.

5.3 Learning Preferences

A safe and engaging space is conducive to learning. However, this may not be found outside traditional classrooms where the possibility of miscommunication has a larger impact and imaginably inhibits students from using FL. Research suggests that practicing a foreign language with a desktop or tablet-based CALL applications can reduce anxiety (Golonka et al. 2014; Timpe-Laughlin, Sydorenko, and Daurio 2020; Bashori et al. 2020). We see this to hold true in CILLE too. As seen in fig. 12, participants noted that, as compared with humans, speaking with AI Agents was generally more comfortable (Figure 12a) and less anxiety-ridden (Figure 12b) without a loss in engagement (Figure 12c). Thus, CILLE uniquely combines the comfort of practicing with technology with the authenticity of the real-world via immersion.

Most students indicated that they prefer the CILLE over other widely available technology-based methods such as watching videos. A student noted, "It felt like I was actually talking to someone". This indicates interest in the student community to communicate and learn with AI Agents, a finding supported by (Gallacher, Thompson, and Howarth 2018).

6 Study Limitations

While the number of participants was few, it allowed for a thorough mixed-methods study with a focus on individual differences of interaction patterns, preferences, and learning gains. Our participants knew how to read Pinyin and had conceptual knowledge of tones prior to the study. Their enrollment in the university-provided Chinese course was an indication of interest and motivation, two factors known to influence success and not discussed in this paper.

External exposure from the classroom or otherwise could have influenced certain aspects of proficiency such as paying attention to tones while listening or speaking. To mitigate some effects, we intentionally chose content that would not be covered in the Chinese course that students were enrolled in. The overlap between exposure to external content and the experiment's content was seen to be zero to minimal through students' diary entries.

While comparing with other technologies, the effect of novelty and lack of control for alternative variables are limitations (a comparable technology was not available for control). However, this survey is not used to determine which is better than what; rather only to elicit students' interest in learning with AI Agents in comparison to others.

The effect of novelty on the quality of interactions was not deterministically measured. We assume that prolonged, repeated exposure combined with onboarding sessions and time with technology outside the learning sessions limits such effects. We observed students expressing frustration with technology bugs and saw them feel tired after consuming the learning material; both point towards a limited effect of novelty on educational benefits.

7 Conclusion and Future Work

In this paper, we demonstrated and evaluated CILLIE, forward-thinking method for learning a foreign language in which we paired multiple AI Agents with an XR environment in which students can interact multimodally and collaboratively.

CILLIE opens up a window for teaching conversational foreign language at discourse level with multimodal, multi-party input, immersed in XR. We saw that XR brought visual context to the students while the AI provided opportunities to roleplay conversations inside the visual context. Our XR could immerse students in a way that did not require markered devices and allowed long-term collaboration between students. Our AI understood multimodal input i.e. it could "see", "hear", "speak" and "show" students conversational aspects related to street market shopping in China. The AI Agents knew when they were being spoken to (without a wake-word), when to respond (irrespective of being addressed) and how to respond in a simultaneous multi-party dialogue; creating a naturalistic conversational/social immersion.

Overall, the interactions were used as expected and lead to a statistical improvement in vocabulary, listening, comprehension, and speaking scores in a user study with 10 participants. The interactions, according to the participating students, were fun, engaging, and found to create a generally more comfortable and less anxiety-ridden space for them to practice using the target language.

The findings from this experiment break ground for new research: creating a new kind of classroom where students, AI agents, and educators work in synergy. It leads to many benefits for the FL students, namely: improvements in students' willingness to communicate in the target language, culture absorption, incidental learning, learning out of need, and other benefits of actual in-country immersion programs without the hassle.

Further, we motivate research in immersive AI for education. Multimodal conversational AI research is in a nascent stage and presents two challenges. First, creating content requires educators to rely on technical experts who themselves will encounter hard research challenges. Eventually, we foresee the ability for rapid content generation by educators and new automated AI tools. Second, multimodal, multi-party interactions with AI agents can be greatly improved in various individual aspects such as recognition of non-native speech (Ubale, Qian, and Evanini 2018) as well as higher-level interpretation of what is happening in the room (including classroom dynamics) and how to respond to it. We see AI taking on roles beyond conversational opportunities such as providing motivation, instruction, and feedback for language learning. To better convey students from the real world into a virtual one, further research in design methods is also foreseen. More intelligent, realistic agents capable of performing several expressions, gestures, and movements in an easily deployable/affordable XR space is a challenge that calls for interdisciplinary research with education as its guiding light.

Acknowledgments

The authors would like to thank Dr. Michael Wei, professor at the University of Missouri-Kansas City, for sharing his expertise in statistical analysis and for helping with user speaking test scoring. The authors would also like to thank RPI and IBM for funding this project as a part of their collaboration through the AI Horizons Network (AIHN)

Appendices

A Sample Questions used in Tests

A.1 Vocabulary Test

A.1.1 Category 1: Hanzi options

Question: Which of the following means “money” in Chinese?

Options: (a) 钱 (b) 千 (c) 买 (d) 金

A.1.2 Category 2: Pinyin options

Question: What is the correct translation for “banana”?

Options: (a) píng guǒ (b) xiāng jiāo (c) píng ān (d) luó bo

A.2 Listening and Transcription

A.2.1 Category 1: Hanzi Options

Audio clip plays: 这个西瓜很甜。(Meaning, this watermelon is very sweet)

Options: (a) 那个苹果很甜。(b) 那个橙子很酸。 (c) 这个西瓜很甜。(d) 这个西瓜很贵。

A.2.2 Category 2: Pinyin Options

Audio clip plays: 苹果很好吃 (Meaning, the apple is delicious)

Options: (a) Cǎo méi yǒu diǎn suān. (b) Cǎo méi hěn hǎo kàn. (c) Píng ān hěn hǎo chī. (d) Píng guǒ hěn hǎo chī.

A.3 Listening Comprehension

A.3.1 Category 1: Word Level Comprehension

Audio clip plays: 酸 (Meaning, sour)

Options: (a) Delicious (b) Sweet (c) Sour (d) Free

A.3.2 Category 2: Conversation Level Comprehension

Audio clip plays:

Woman - 你有没有菠萝? (Do you have pineapple?)

Man - 我有菠萝. 我的菠萝很甜 (I have pineapple. Mine are very sweet)

Woman - 多少钱? (How much is it?)

Man - 十块 (10 yuan)

Woman - 我买你的菠萝 (I will buy your pineapple)

Question: What is the woman trying to do？

Options:

(a) She is trying to buy pineapples.

(b) She is trying to sell pineapples.

(c) She is trying to get a discount.

(d) She is trying to open a store.

A.4 Conversation

A.4.1 Part 1: Interview

Interviewer: If you want to buy apples，how would you ask the price in Chinese？

Student (sample answer): 这个苹果多少钱？ (How much is this apple?)

A.4.2 Part 2: Roleplay

Interviewer (as a shopkeeper): 来看看我的橙子，很甜很好吃。 (Come and see my oranges， they are sweet and delicious）

Student (as a buyer): 这个多少钱？(How much is this?)

Interviewer (as a shopkeeper): 苹果6块，很便宜了 （Apples are 6 kuai, they are cheap）

Student (as a buyer): 能便宜点吗？(can it be cheaper?)

Interviewer (as a shopkeeper): 已经很便宜了，4块，不能再便宜了 (they‘re already cheap， 4 kuai， can’t be cheaper)

Student (as a buyer): 嗯，好的(Ok, ok)

A.4.3 Part 3: Interview based on part 2

Interviewer: 你要买的水果是什么? (what was the fruit you wanted to buy?)

Student: 我要买橙子 (I want to buy oranges)

Interviewer: 你拿到了多少折扣？ (How much discount did you get?)

Student: I don't know

Interviewer: 你付了多少钱？ (How much did you pay?)

Student: 我付8块 (I paid 8 yuan)

References

Adair-Hauck, Bonnie, Eileen W Glisan, Keiko Koda, Elvira B Swender, and Paul Sandrock. 2006. "The Integrated Performance Assessment (IPA): Connecting Assessment to Instruction and Learning." *Foreign Language Annals* 39 (3): 359–82.

Ahmadi, Farida Badri, and Essa Panahandeh. 2016. "The Role of Input-Based and Output-Based Language Teaching in Learning English Phrasal Verbs by Upper-Intermediate Iranian Efl Learners." *Journal of Education and Learning* 10 (1): 22–33.

Anderson, James N, Nancie Davidson, Hazel Morton, and Mervyn A Jack. 2008. "Language Learning with Interactive Virtual Agent Scenarios and Speech Recognition: Lessons Learned." *Computer Animation and Virtual Worlds* 19 (5): 605–19.

Avcı, Şirin Küçük, Ahmet Naci Coklar, and Aslıhan İstanbullu. 2019. "The Effect of Three Dimensional Virtual Environments and Augmented Reality Applications on the Learning Achievement: A Meta-Analysis Study." *Egitim Ve Bilim* 44 (198).

Ayedoun, Emmanuel, Yuki Hayashi, and Kazuhisa Seta. 2015. "A Conversational Agent to Encourage Willingness to Communicate in the Context of English as a Foreign Language." *Procedia Computer Science* 60: 1433–42.

Bashori, Muzakki, Roeland van Hout, Helmer Strik, and Catia Cucchiarini. 2020. "Web-Based Language Learning and Speaking Anxiety." *Computer Assisted Language Learning* 0 (0): 1–32.

Bayser, Maira Gatti de, Melina Alberio Guerra, Paulo Cavalin, and Claudio Pinhanez. 2018. "Specifying and Implementing Multi-Party Conversation Rules with Finite-State-Automata." In *Workshops at the Thirty-Second Aaai Conference on Artificial Intelligence*.

Bayser, Maira Gatti de, Claudio Pinhanez, Heloisa Candello, Marisa Affonso, Mauro Pichiliani Vasconcelos, Melina Alberio Guerra, Paulo Cavalin, and Renan Souza. 2018. "Ravel: A MAS Orchestration Platform for Human-Chatbots Conversations."

Bennett, Milton J. 1997. "How Not to Be a Fluent Fool: Understanding the Cultural Dimension of Language." *New Ways in Teaching Culture*, 16–21.

Bibauw, Serge, Thomas François, and Piet Desmet. 2019. "Discussing with a Computer to Practice a Foreign Language: Research Synthesis and Conceptual Framework of Dialogue-Based Call." *Computer Assisted Language Learning* 32 (8): 827–77.

Bricken, William. 1990. "Learning in Virtual Reality." Human Interface Technology Laboratory.

Brown, John Seely, Allan Collins, and Paul Duguid. 1988. “Situated Cognition and the Culture of Learning.” Bolt Beranek; Newman, Inc.

Canto, Silvia, and Kristi Jauregi Ondarra. 2017. “Language Learning Effects Through the Integration of Synchronous Online Communication: The Case of Video Communication and Second Life.” *Language Learning in Higher Education.* 7 (1): 21–53.

Cheng, Alan, Lei Yang, and Erik Andersen. 2017. “Teaching Language and Culture with a Virtual Reality Game.” In *Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems*, 541–49.

ChengChiang Chen, Julian, and Sarah Kent. 2020. “Task Engagement, Learner Motivation and Avatar Identities of Struggling English Language Learners in the 3D Virtual World.” *System* 88 (February): 102168.

Chien-pang Wang, Yu-Ju Lan, Wen-Ta Tseng, Yen-Ting R. Lin & Kao Chia-Ling Gupta. 2019. “On the effects of 3D virtual worlds in language learning – a meta-analysis”. Computer Assisted Language Learning.

Choudhury, Murshed Haider. 2013. “Teaching Culture in Efl: Implications, Challenges and Strategies.” *IOSR Journal of Humanities and Social Science* 13 (1): 20–24.

Coniam, David. 2008. “Evaluating the Language Resources of Chatbots for Their Potential in English as a Second Language.” *ReCALL* 20 (1): 98–116.

Crowe, Aaron. n.d. “Haggling Is a Lost Art in the U.s. - or Is It Just Evolving? - Aol Finance.” https://www.aol.com/article/finance/2014/10/06/haggling-lost-art-america-or-evolving/20971447/?guccounter=1.

Cruz-Neira, Carolina, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. 1992. “The Cave: Audio Visual Experience Automatic Virtual Environment.” *Communications of the ACM* 35 (6): 64–73.

Divekar, Rahul R, Jaimie Drozdal, Yalun Zhou, Ziyi Song, David Allen, Robert Rouhani, Rui Zhao, Shuyue Zheng, Lilit Balagyozyan, and Hui Su. 2018. “Interaction Challenges in Ai Equipped Environments Built to Teach Foreign Languages Through Dialogue and Task-Completion.” In *Proceedings of the 2018 Designing Interactive Systems Conference*, 597–609. ACM.

Divekar, Rahul R, Jeffrey O Kephart, Xiangyang Mou, Lisha Chen, and Hui Su. 2019. “You Talkin’to Me? A Practical Attention-Aware Embodied Agent.” In *Human-Computer Interaction–Interact*. Vol. 2019.

Divekar, Rahul R, Xiangyang Mou, Lisha Chen, Maira Gatti de Bayser, Melina Alberio Guerra, and Hui Su. 2019. “Embodied Conversational Ai Agents in a Multi-Modal Multi-Agent Competitive Dialogue.” In. IJCAI.

Divekar, Rahul R, Matthew Peveler, Robert Rouhani, Rui Zhao, Jeffrey O Kephart, David Allen, Kang Wang, Qiang Ji, and Hui Su. 2018. “CIRA: An Architecture for Building Configurable Immersive Smart-Rooms.” In *Proceedings of Sai Intelligent Systems Conference*, 76–95. Springer.

Divekar, Rahul R, Yalun Zhou, David Allen, Jaimie Drozdal, and Hui Su. 2018. “Building Human-Scale Intelligent Immersive Spaces for Foreign Language Learning.” *iLRN 2018 Montana*, 94.

Doremalen, Joost van, Lou Boves, Jozef Colpaert, Catia Cucchiarini, and Helmer Strik. 2016. “Evaluating Automatic Speech Recognition-Based Language Learning Systems: A Case Study.” *Computer Assisted Language Learning* 29 (4): 833–51.

Dörnyei, Zoltán. 2009. *The Psychology of Second Language Acquisition*. Oxford University Press Oxford.

Edwards, Chad, Edwards, Autumn, Spence, Patric R., Lin, Xialing. 2018. “I, teacher: using artificial intelligence (AI) and social robots in communication and instruction.” *Communication Education* 67 (4): 473-480.

Flores, Jorge Francisco Figueroa. 2015. “Using Gamification to Enhance Second Language Learning.” *Digital Education Review*, no. 27: 32–54.

Freed, Barbara F, Norman Segalowitz, and Dan P Dewey. 2004. “Context of Learning and Second Language Fluency in French: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs.” *Studies in Second Language Acquisition* 26 (2): 275–301.

Fryer, L. K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. Language Learning & Technology, 24(2), 8–22.

Fryer, Luke, and Rollo Carpenter. 2006. “Bots as Language Learning Tools.” Language Learning & Technology 10 (3): 8–14.

Gallacher, Andrew, Andrew Thompson, and Mark Howarth. 2018. “‘My Robot Is an Idiot!’–Students’ Perceptions of AI in the L2 Classroom.” *Future-Proof CALL: Language Learning as Exploration and Encounters–Short Papers from EUROCALL 2018*, 70.

Golonka, Ewa M., Anita R. Bowles, Victor M. Frank, Dorna L. Richardson, and Suzanne Freynik. 2014. “Technologies for Foreign Language Learning: A Review of Technology Types and Their Effectiveness.” *Computer Assisted Language Learning* 27 (1): 70–105.

Güzel, Serhat, and Selami Aydin. 2016. “The Effect of Second Life on Speaking Achievement.” In *4th Global J. Of Foreign Lang. Teaching*, 6:236–45. 4.

Hassani, Kaveh, Ali Nahvi, and Ali Ahmadi. 2016. “Design and Implementation of an Intelligent Virtual Environment for Improving Speaking and Listening Skills.” *Interactive Learning Environments* 24 (1): 252–71.

Hsiao, Tina, and Jules Kay. n.d. “How to Bargain: The Ultimate Guide to Scoring Deals in the Markets of Asia.” Cable News Network. http://travel.cnn.com/ultimate-guide-bargaining-asia-896226/.

Hubbard, Philip. 2006. “Evaluating Call Software.” *Calling on CALL: From Theory and Research to New Directions in Foreign Language Teaching*, 313–38.

Hung, Hsiu-Ting, Jie Chi Yang, Gwo-Jen Hwang, Hui-Chun Chu, and Chun-Chieh Wang. 2018. “A Scoping Review of Research on Digital Game-Based Language Learning.” *Computers & Education* 126: 89–104.

Kim, Na-Young, Yoonjung Cha, and Hea-Suk Kim. 2019. “Future English Learning: Chatbots and Artificial Intelligence.” *Multimedia-Assisted Language Learning* 22 (3): 32–53.

Kot Artunç, Esma, and Deniz Ortaçtepe Hart. 2020. “Interactional Competence in Paired Speaking Tests: A Study on Proficiency-Based Pairings.” System 89 (April): 102194.

Lai, Chun, and Guofang Li. 2011. “Technology and Task-Based Language Teaching: A Critical Review.” CALICO Journal 28 (2): 498–521.

Lave, Jean, and Etienne Wenger. 1991. *Situated Learning*. New York, NY, USA: Cambridge Univ. Press.

Legault, Jennifer, Jiayan Zhao, Ying-An Chi, Weitao Chen, Alexander Klippel, and Ping Li. 2019. “Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning.” *Languages* 4 (1): 13.

Li, Ping, Sara Sepanski, and Xiaowei Zhao. 2006. “Language History Questionnaire: A Web-Based Interface for Bilingual Research.” *Behavior Research Methods* 38 (2): 202–10.

Lin, James P. 2013. *Modern Chinese Textbook 1A*. 2nd ed. Better Chinese.

Morton Hazel, and Mervyn Jack. “Scenario-Based Spoken Interaction with Virtual Agents”. *Computer Assisted Language Learning* 18 (3): 171–191.

Morton, Hazel, Nancie Gunson, and Mervyn Jack. 2012. “Interactive Language Learning Through Speech-Enabled Virtual Scenarios.” *Advances in Human-Computer Interaction* 2012.

O'Brien, Mary, Levy, Richard, and Orich, Annika. 2009. “Virtual Immersion: The Role of CAVE and PC Technology.” *CALICO Journal, 26*(2).

Oxford, Rebecca, and Christina Gkonou. 2018. “Interwoven: Culture, Language, and Learning Strategies.” *Studies in Second Language Learning and Teaching* 8 (June): 403–26.

Radianti, Jaziar, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. “A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda.” *Computers and Education* 147: 103778.

Reuland, Daniel S, Lisa M Slatt, Marco A Alemán, Alicia Fernandez, and Darren DeWalt. 2012. “Effect of Spanish Language Immersion Rotations on Medical Student Spanish Fluency.” *Family Medicine-Kansas City* 44 (2): 110.

Sercu, Lies. 2010. “Assessing Intercultural Competence: More Questions Than Answers.” *Testing the Untestable in Language Education* 1734.

Sharda, Ramesh, Nicholas C. Romano, Joyce A. Lucca, Mark Weiser, George Scheets, Jong-Moon Chung, and Catherine M. Sleezer. 2004. “Foundation for the Study of Computer-Supported Collaborative Learning Requiring Immersive Presence.” *J. Manage. Inf. Syst.* 20 (4): 31–63.

Shawar, Bayan Abu. 2017. “Integrating Call Systems with Chatbots as Conversational Partners.” *Computación Y Sistemas* 21 (4): 615–26.

Steves, Rick. n.d. “Top Tips for Bargaining at Europe’s Markets | Smartertravel.” https://www.smartertravel.com/top-tips-for-bargaining-at-europes-markets/.

Taguchi, Naoko. 2015. “‘Contextually’ Speaking: A Survey of Pragmatic Learning Abroad, in Class, and Online.” System 48 (February): 3–20.

Timpe-Laughlin, Veronika, Tetyana Sydorenko, and Phoebe Daurio. 2020. “Using Spoken Dialogue Technology for L2 Speaking Practice: What Do Teachers Think?” *Computer Assisted Language Learning* 0 (0): 1–24.

Tsai, Yu-Ling, and Chin-Chung Tsai. 2018. “Digital Game-Based Second-Language Vocabulary Learning and Conditions of Research Designs: A Meta-Analysis Study.” *Computers & Education* 125: 345–57.

Ubale, Rutuja, Yao Qian, and Keelan Evanini. 2018. “Exploring End-to-End Attention-Based Neural Networks for Native Language Identification.” In *2018 Ieee Spoken Language Technology Workshop (Slt)*, 84–91. IEEE.

Wang, Yanjun, Scott Grant, and Matthew Grist. 2020. “Enhancing the Learning of Multi-Level Undergraduate Chinese Language with a 3D Immersive Experience - An Exploratory Study.” Computer Assisted Language Learning, July, 1–19.

Tang, Xiaofei, and Naoko Taguchi. 2020. "Designing and Using a Scenario-Based Digital Game to Teach Chinese Formulaic Expressions." *Calico Journal* 37.1.

Yamazaki, Kasumi. 2018. “Computer-assisted learning of communication (CALC): A case study of japanese learning in a 3D virtual world.” *ReCALL - The Journal of EUROCALL*, 30(2), 214-231.

Yang, M., and W. Liao. 2014. “Computer-Assisted Culture Learning in an Online Augmented Reality Environment Based on Free-Hand Gesture Interaction.” *IEEE Transactions on Learning Technologies* 7 (2): 107–17.

Zhao, Rui, Kang Wang, Rahul Divekar, Robert Rouhani, Hui Su, and Qiang Ji. 2018. “An Immersive System with Multi-Modal Human-Computer Interaction.” In *2018 13th Ieee International Conference on Automatic Face & Gesture Recognition (Fg 2018)*, 517–24. IEEE.