Article

Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning

Jennifer Legault *[ID]**, Jiayan Zhao**[ID]**, Ying-An Chi, Weitao Chen, Alexander Klippel and Ping Li ***

Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA 16802, USA; juz64@psu.edu (J.Z.); yqc5514@psu.edu (Y.-A.C.); wxc5182@psu.edu (W.C.); klippel@psu.edu (A.K.)

* Correspondence: jtl190@psu.edu (J.L.); pul8@psu.edu (P.L.)

check for updates

Abstract: Learning a second language (L2) presents a significant challenge to many people in adulthood. Platforms for effective L2 instruction have been developed in both academia and the industry. While real-life (RL) immersion is often lauded as a particularly effective L2 learning platform, little is known about the features of immersive contexts that contribute to the L2 learning process. Immersive virtual reality (iVR) offers a flexible platform to simulate an RL immersive learning situation, while allowing the researcher to have tight experimental control for stimulus delivery and learner interaction with the environment. Using a mixed counterbalanced design, the current study examines individual differences in L2 performance during learning of 60 Mandarin Chinese words across two learning sessions, with each participant learning 30 words in iVR and 30 words via word–word (WW) paired association. Behavioral performance was collected immediately after L2 learning via an alternative forced-choice recognition task. Our results indicate a main effect of L2 learning context, such that accuracy on trials learned via iVR was significantly higher as compared to trials learned in the WW condition. These effects are reflected especially in the differential effects of learning contexts, in that less successful learners show a significant benefit of iVR instruction as compared to WW, whereas successful learners do not show a significant benefit of either learning condition. Our findings have broad implications for L2 education, particularly for those who struggle in learning an L2.

Keywords: immersive virtual reality; second language acquisition; effects of learning context; embodied cognition

1. Introduction

Advances in technology have led to an increasingly connected world, bringing forth an expanding population of bilinguals who speak two or more languages. Many countries require regular use of at least two languages, with many children successfully learning both languages simultaneously or sequentially at an early age. However, learning a second language during adulthood has been consistently shown to be markedly difficult.

While previous assumptions of critical or sensitive periods for language learning remain debated (Lenneberg 1967; Johnson and Newport 1989), current behavioral and neuroimaging findings indicate that the fundamental principles of learning for child native language (L1) and adult second language (L2) acquisition may be similar but the methods and contexts that support children and adults in language learning vary greatly, leading to significant differences between child L1 and adult L2 (Hernandez and Li 2007; Hernandez et al. 2005; MacWhinney 2012). For example, the context in which an L2 is learned has been shown to influence L2 learning performance. Studies comparing real-life (RL) immersive L2 learning (moving to a foreign country to learn its language) have shown reduced interferences from L1 to L2 for the RL context versus typical classroom instruction. Linck et al. (2009)

examined cross-linguistic interference in English monolinguals who learned L2 vocabulary either in Spain via study abroad or through classroom instruction in the United States. They found significantly reduced L1–L2 interference in students who learned via study abroad, as well as higher L2 proficiency in these individuals, as compared to those who learned in the classroom.

However, RL immersive L2 learning is not always possible for or accessible to every student, since many do not have the resources or time to move to another country merely to learn a second language. Furthermore, it is difficult to experimentally manipulate or examine the specific contributions of RL immersive L2 learning since no two people studying abroad will learn the same items in exactly the same manner. In fact, many experiences abroad will vary significantly depending on the region, environment, and people the L2 learners surround themselves with. Therefore, examining the efficacy of RL immersive L2 learning can be challenging and difficult to replicate. Given these challenges, a growing area of research has focused on virtual environments and more recently immersive virtual reality as experimentally versatile L2 learning platforms.

Recent advances in technology have led to the availability of immersive virtual reality (iVR) environments that allow users access to a 360-degree field of regard and access to the environment using a head-mounted display. This technology can provide accessible environments that are engaging, highly immersive, and interactive. Previous literature has indicated the effectiveness of L2 learning platforms via simulated action videos (Jeong et al. 2010) and online computer-based virtual environments (Berns et al. 2013; Chen 2016; Ibáñez et al. 2011; Lan et al. 2014; Levak and Son 2017; Si 2015; Wang et al. 2017). Uniquely, this study examines performance in adults learning Mandarin Chinese vocabulary through iVR versus explicit L1 to L2 word–word (WW) paired association learning. The current study further compares two different iVR environments with regard to their effectiveness, one focused on enabling a high degree of interaction and another on enabling a high degree of spatial navigation during L2 learning; both were compared with L1–L2 word–word paired association learning.

1.1. L2 Learning Context Affects Learning Performance

Given the difficulty most adults exhibit when learning an L2, a large body of research has focused on identifying the effects of specific aspects, including learning context, on L2 learning success. A comprehensive review by Collentine and Freed (2004) overviewed conditions by which studying abroad, domestic immersive learning, and at-home (in a classroom) learning were associated with L2 success. Participants who study abroad have been shown to have improved L2 fluency (Freed et al. 2004), enhanced communication skills (Lafford 1995), and excel at L2 vocabulary acquisition (Milton and Meara 1995), with the caveats that not all studies directly compared at home learning with studying abroad, and not all participants benefited equally from study abroad instruction.

This inconsistency in finding L2 context-related differences has led to examination of individual differences related to L2 acquisition success in order to examine which individuals may gain the greatest benefit from L2 instruction. Participants with a stronger foundation in grammar and reading skills in the L2 have been shown to benefit more from the study abroad experience over novice learners (Brecht et al. 1995). For example, Brecht et al. (1995) examined oral proficiency scores in participants before and after L2 RL immersion, showing that individuals who had better L2 reading and grammatical skills showed greater oral proficiency gains as compared to those who started off with lower L2 reading and grammar proficiency. Meanwhile, other studies have concentrated on the link between L1 proficiency and L2 proficiency, which seem to be consistently interdependent (Cummins 1991). Other posited predictors of L2 learning success include individual differences in cognitive control (Abutalebi and Green 2007), including working memory (Linck et al. 2014; Miyake and Friedman 1998) and conflict monitoring (Abutalebi et al. 2012). However, whether these predictors vary as a function of L2 learning context warrants additional research.

L2 learning context effects may be partially driven by the different mechanisms by which L1 and L2 instruction typically occur. Immersive learning is hypothesized to reflect L1 acquisition by promoting embodied and perceptually-relevant mental representation of learned L2 items. By contrast, classroom

instruction is often removed from embodied and perceptual–sensorimotor experiences. In line with this reasoning, we link L2 learning context effects to the cognitive theory of embodied representation, suggesting that a key element that underlies the success of immersive environments in L2 instruction is the level of embodiment they can provide (see Lan et al. 2015, 2014). Cognitive psychology has provided a framework in the last two decades on embodied representation, particularly with regard to how representation is based on visual, spatial, auditory, and other modality-specific experiences from the environment (Fischer and Zwaan 2008; Barsalou 2008; Lan et al. 2015). For the purposes of the current study, our definition of embodiment assumes that body-specific and modality-specific (e.g., auditory, visual, tactile) experiences are involved in the learner's integral representation of concepts, objects, and action. This includes all bodily sensation and perception of the input properties/features, which, when engaged and used, would recruit neural regions responsible for perception and sensorimotor actions (Aziz-Zadeh and Damasio 2008; Barsalou et al. 2003; Glenberg et al. 2008; Willems and Casasanto 2011).

Learning an L1 is thought to incur rich perceptual and sensorimotor experiences with features of objects (e.g., shape, size, color, orientation, location), and with relations between the objects in the space, along with interpersonal interactions. For example, babies begin to learn their first set of words in the context of crawling in the house and interacting with the parents. By contrast, adult classroom L2 instruction often involves rote memorization and direct translation of L2 to L1, detached from the space in which language use typically takes place. The Unified Competition Model (UCM; MacWhinney 2012) argues that such differences between how we instruct L1 and L2 underscore the importance of perception–action enabled context for successful child L1 vs. the less successful adult L2 learning, due to the use (or nonuse) of embodied, immersive, environments. Furthermore, typical classroom L2 learning lacks the relevant perceptual–visual–sensory features in the representation of the vocabulary (e.g., shape, size, motion, and location of an item to be learned). In contrast, immersive learning can engage the learner in a natural, embodied, perception–action rich context.

Finally, many studies now emphasize the role of individual differences in L1 performance and cognitive ability in mediating L2 learning performance. Studies have found that L2 learning performance can be related with greater working memory and inhibitory control abilities (Linck et al. 2009). These cognitive functions are posited to contribute to L2 learning by decreasing the level of cross-linguistic activation. Studies have also shown that L2 experience is related to greater conflict monitoring ability (Abutalebi et al. 2012). Further, as introduced above, increased L1 proficiency before L2 learning is often associated with great L2 learning gains (Brecht et al. 1995). In a previous study (Legault et al. 2018) examining the effects of L2 learning context across an online virtual environment (VE) versus picture–word (PW) paired association conditions, we found that working memory was associated with L2 performance in VE but not PW learners. Therefore, the current study will examine individual differences in working memory, conflict monitoring, and L1 proficiency and their possible relationships with L2 learning. Since our iVR environments require our participants to navigate the VR environment while they learn, we will also examine the relationship between spatial navigation ability and L2 performance. Finally, we will examine whether the relationships between these cognitive factors and L2 performance may be mediated by L2 learning context.

1.2. Virtual Environments Can Provide Effective and Engaging Learning Experiences

Virtual environments (VEs) can be broadly defined as 3D interactive environments that are computer-generated. VEs are widely used across various technologies including VR, augmented reality (AR), mixed reality (MR), virtual habitats presented via desktops, and some 3D videos and video games. For example, MR blends the VE with a real environment. Immersive VR is defined by its 360 degree field of regard, where users can see a different part of the virtual environment by turning their head and body to look in any direction). Immersive VR contexts enable users to manipulate or interact with the environment seamlessly as in real-life situations (Costello 1997; Jerald 2016; Milgram and Kishino 1994). Immersive VR and MR technology provide the greatest amount of immersion and interaction amongst VE platforms.

While many of these VE platforms have been traditionally used for entertainment purposes, an upsurge of schools, hospital settings, and research institutions have examined their ability to provide active learning environments, leading to the existence of virtual learning environments (VLEs). These VLEs have been found to promote motor function during stroke rehabilitation as compared to conventional therapy (see Lohse et al. 2014), to aid in sensorimotor training and rehabilitation (Adamovich et al. 2009), to provide an effective spatial navigation learning platform for some amnesic patients (Brooks et al. 1999), and to produce effective instructions of science courses (Johnson-Glenberg et al. 2014) and L2 learning (Jeong et al. 2010; Lan et al. 2014).

There is a sizeable literature on the use of VEs for L2 language instruction; for an overview and comparison of studies, see Table 1. In this table, the level of immersion was defined by the display device(s) used, where projection and headsets were rated as high immersion, 3D spaces in desktops were rated as medium immersion, and 2D videos or platforms were categorized as low immersion. The level of interaction depended on whether the actions were conducted via 3D handsets or body movements via Kinect (high), via computer mouse (medium), or if there were no actions available (none). The vast majority of these studies did not use comparison groups since a main goal of the studies were to examine whether VEs could be effectively used to teach an L2, and to examine the aspects of VEs that enabled effective L2 learning. However, three studies did focus on the effect of learning context by comparing VE L2 learning to text-based, picture–word paired-association or classroom-based learning (Jeong et al. 2010; Lan et al. 2014; Si 2015), which we briefly review below.

Jeong et al. (2010) compared text-based L2 learning with situation-based learning, where all participants learned through video watching. In the text-based learning condition, participants viewed a person holding up a piece of paper with an L1 item written in the center along with an auditory presentation of the L2 word. In the situation-based condition, participants viewed a person (or several people) performing some type of task along with the corresponding L2 auditory word. This condition was used to simulate an immersed environment. They found a behavioral advantage for the situation-based learning group as compared to the text-based learning group. Furthermore, they examined neural activation during testing of these target items and found that participants who learned via the situation-based condition recruited neural regions implicated in embodied cognition networks during the task (the middle frontal gyrus and inferior parietal lobe), whereas the text-based learning group did not.

Lan et al. (2014) examined the effect of L2 learning context, comparing picture–word (PW) association vocabulary learning versus learning via an online VE platform. The PW association learning involved participants seeing a picture on the screen while hearing the corresponding auditory Chinese word. The VE learning used *SecondLife*, a popular online gaming and social network platform that allowed participants to create their own avatar, move in a 3D space, and click on items to hear their corresponding Chinese word. Results from the study indicated that the VE group showed a faster learning acceleration as compared to the PW group, and needed only half the number of exposures as the PW group (measured by number of clicks on items to hear their corresponding Chinese word) to attain the same level of accuracy. These findings indicate that VEs can provide effective learning platforms for L2 learning, sometimes not directly in accuracy but in learning speed. A structural magnetic resonance imaging (sMRI) study examined brain structure changes in these same individuals, and found that participants who trained in the VE context showed that L2 performance was correlated with the anterior cingulate cortex, inferior frontal lobe (IFG), and inferior parietal lobe (IPL), neural regions involved in embodied networks (Legault et al. 2018). Participants who learned in the PW environment recruited a less distributed network. Further, L2 learning performance correlated with different neural structures depending on the L2 learning context used. The PW group showed their L2 learning performance was associated with cortical thickness (CT) in the right IFG, a region hypothesized to be associated with explicit L2 instruction (Stein et al. 2014). The learning performance for the VE group was alternatively correlated with CT in the right IPL, a region shown to be associated with highly effective L2 learning (Richardson and Price 2009). Together, these findings indicate that L2 learning context can affect both L2 learning performance and subsequent brain changes.

Table 1. Summary of studies examining the efficacy of virtual environment (VE) platforms for L2 instruction.

Study	Participants	L2 Training Methods	VE Immersion	VE Interaction	Main Behavioral Findings
Berns et al. (2013)	85 Spanish speakers who had learned German for 8 months prior to VE training	L2 Training across five sessions for a total of 60–95 min. L2 testing: Vocabulary pre-test and multiple choice post-test	• Level: Medium • Display: Desktop 3D platform • Platform: VirtUAM and OpenSim	• Level: Medium • Functions: computer mouse, keyboard, avatar-based movement and text-chat	• Participants significantly improved in L2 writing, reading, and listening skills after VE training. • Effective VE features: training with game-like applications lead to highly motivated, engaged, and facilitated L2 learning.
Chen (2016)	9 participants aged between 21 and 55 from across Europe, Asia, and Africa learning English as a foreign language	L2 training across five weeks consisting of ten 90-min sessions. L2 testing: learning journal and interview questions on VE efficacy	• Level: Medium • Display: Desktop 3D platform • Platform: Second Life	• Level: Medium • Functions: computer mouse, keyboard, avatar-based movement and flexible chat logs	VE: promoted collaboration and L2 communication skills. Effective VE features: • Flexible chat modes • Build option promoted collaboration • Engaging platform enabled fun learning • Multicultural aspect of program enhanced learning experience
Ibáñez et al. (2011)	12 non-native Spanish learners and six foreign language teachers	Spanish learners and teachers were grouped in six different VEs. L2 testing: interview questions on VE efficacy.	• Level: Medium • Display: Desktop 3D platform • Platform: Open Wonderland	• Level: Medium • Functions: computer mouse, keyboard, avatar-based movement, chatbots and non-player characters (NPCs)	VE: engaging L2 platform. Effective VE features: natural chatting via NPCs and chatbots led to effective communication skills.
Jeong et al. (2010)	31 Native Japanese speakers learning Korean (mean age: 21.6)	One 3-h session to learn 12 items via situation-based videos and 12 items via text-based videos. L2 testing: situation-based and text-based contexts.	• Level: Low • Display: Desktop videos	• Level: None • Functions: None; passive video watching	Situation-based items were remembered more flexibly than text-based items. • Text-based item accuracy was dependent on testing context • Situation-based items were remembered regardless of context
Lan et al. (2014)	36 Native English speakers learning Chinese: • 17 picture–word (PW) L2 learners (Mean age = 21.29) • 19 VE L2 learners (Mean age = 22.21)	Seven 1-h training sessions on 90 vocabulary words. L2 testing: four alternative-forced choice (4AFC) recognition of all items.	• Level: Medium • Display: Desktop 3D platform • Platform: Second Life	• Level: Medium • Functions: computer mouse, keyboard, and avatar-based movement	VE: more effective than PW learning. • VE: faster L2 learning acceleration than the PW group • VE group needed half the exposures to reach same accuracy as PW group • Caveat: without controlling exposures, PW group performed with higher accuracy than VE group

Table 1. *Cont.*

Study	Participants	L2 Training Methods	VE Immersion	VE Interaction	Main Behavioral Findings
Levak and Son (2017)	35 students (Age: 18–30) • 18 English speakers learning Croatian • 17 Croatian speakers learning English • L2 conditions: (a) Second Life; (b) Skype; (c) Second Life + Skype	L2 course—supplement sessions contained eight communication-based tasks. English and Croatian participants were paired together. L2 testing: listening comprehension	• Level: Medium • Display: Desktop 3D platform • Platform: Second Life	• Level: Medium • Functions: computer mouse, keyboard, avatar-based movement and text-chat	Participants benefited from both Skype and VE sessions. Effective VE features: • Ability to view facial expressions in Skype facilitated L2 listening comprehension • Ability to have an engaging 3D platform and text-chat facilitated L2 listening comprehension
Si (2015)	20 Native English children between the ages of 6 and 8 learning Mandarin Chinese.	One month of VE training, with two 30-min sessions per week using a desktop computer game and Kinect motion sensors	• Level: Medium • Display: Desktop 3D platform • Platform: Unity	• Level: High • Functions: computer mouse and keyboard, avatar-based movement and chat features	VE: increased engagement, L2 vocabulary, and speaking skills Effective VE features: • The use of Kinect for motions to promote embodied experience • Chat feature improved communication skills
Wang et al. (2017)	80 Native Chinese speakers learning English • VE alone • VE + chatbox • VE + time machine • VE + chatbox and time machine	Participants communicated with other characters. L2 testing: Igroup Presence Questionnaire and Immersive Tendencies Questionnaire	• Level: Medium • Display: Desktop 3D platform • Platform: OpenSimulator	• Level: Medium • Functions: computer mouse, keyboard, avatar-based movement and communication features	The participants using the combined VE + chatbox and time machine functions for L2 learning scored significantly higher in perception of immersion and presence as compared to other L2 learning contexts.

Abbreviations: L2: second language; VE: virtual environment.

Si (2015) compared training in a VE versus prior classroom instruction in children. In this study, child participants were raised by parents or grandparents who spoke Chinese but never had any formal training. They underwent classroom L2 instruction for three months, consisting of two sessions per week lasting approximately 30 min each. Next, participants underwent one month of VE-based L2 instruction where they could use their own bodies to move the desktop-based avatars on screen through the use of Kinect technology. Further, chat functions enabled participants to speak with other users in real time. This VE-based learning led to increased engagement of the students, about 20% increase in vocabulary skills, and significant improvement in speaking and pronunciation skills in the majority of students. However, it is difficult to parse these effects of VE as being due to the learning context and not necessarily due to increased practice since the classroom L2 instruction preceded the VE learning (and there were no other learning group comparisons). Furthermore, many of the items and lessons that were learned in the VE learning were repeat items that were first learned during classroom L2 instruction.

A large proportion of studies have promoted the use of VEs with regard to the high degree of communication opportunities, which encourage participant to discuss with other users or teachers in real time via voice chat or text chat, based on user preference, which is proposed to facilitate L2 learning and collaboration (Chen 2016; Ibáñez et al. 2011; Levak and Son 2017; Si 2015; Wang et al. 2017). These studies all found that use of technology-enabled real-time communication facilitated participants' L2 performance and engagement. For instance, Wang et al. (2017) compared a total of 80 participants to examine L2 learning. They divided participants into four groups of 20 participants each: VE alone, VE with chatbox functions (enabling participants to communicate with other users), VE with time machine (which enabled participants to travel to different regions across time), and VE with both chatbox and time machine. Their data indicated that this last condition led to particularly high perceptions of immersion and presence in participants. These findings support the concept that the degree of interaction on a VE platform can affect L2 learning experience.

Ibáñez et al. (2011) conducted a preliminary qualitative investigation on non-native Spanish L2 learners who collaboratively learned Spanish in six different learning environments. Students feedback indicated that participants found the VE platform to be engaging and the communication and collaborative features of the VE to be particularly helpful in understanding how to perform the tasks. Additionally, many of these studies emphasized the importance of avatar-based movements and tools that allow participants to move avatars using bodily movements, and these movements could enhance embodied experiences and 3D game-like environments to promote engagement (Berns et al. 2013; Chen 2016; Ibáñez et al. 2011; Levak and Son 2017; Si 2015; Wang et al. 2017).

Finally, a majority of these studies were conducted on experienced, moderately proficient L2 learners, with students concurrently learning on VEs and taking university language courses due to the foreign language requirements (Levak and Son 2017), or with students having taken courses for 3 months (Si 2015) or 8 months before VE training (Berns et al. 2013). This is likely due to the platforms used, which often required some foundational knowledge in the L2 to perform the tasks. However, some studies included students with varying expertise, from beginner to upper-intermediate level (Chen 2016); some did not report prior L2 experience levels (Ibáñez et al. 2011; Wang et al. 2017). Notably, two studies did examine L2 learning in participants without prior L2 experience in the target L2 since they were geared towards examining L2 learning context effects (Jeong et al. 2010; Lan et al. 2014). A unique aspect of the current study, as compared with all previous work that used desktop 3D VE learning as reviewed, is that we provide the first direct comparison iVR-based L2 learning to WW-based L2 learning.

1.3. Features of Virtual Environment Platforms That Support Learning

While many studies have found VEs provide effective learning platforms, what aspects or features of these VEs lead to learning success is a matter of ongoing investigation. iVR not only provides a platform for student learning but also a useful tool containing various features for researchers

to optimize. First, iVR offers a flexible platform for designing learning contexts that can vary in crucial environmental characteristics (Blascovich et al. 2002; Casasanto and Jasmin 2018); for example, in a real zoo, one cannot easily rearrange animals to different locations to suit a study, while this is a matter of a few clicks when programming a iVR Zoo. At the same time, iVR ensures tight experimental control of stimuli so that environmental features are perceived by every participant in the same way. Second, from a learner's perspective, iVR has the potential to enhance educational outcome because of its ability to simulate a real-life environment in which natural learning takes place (Dede 2009). Third, from a cognitive science perspective, iVR offers a realistic environment to facilitate the construction of *embodied* mental representation, in which interactions can occur naturally. Unlike in classroom-based learning, the iVR learner can integrate visual, auditory, and other sensory features of the learning environment. This is because iVR environments enable participants to move in a realistic 3D space, which is hypothesized to underlie a high sense of embodied presence in a virtual world (Schubert et al. 1999)

An extensive study conducted by Johnson-Glenberg et al. (2014) examined the efficacy of a mixed reality (MR) learning environment in teaching high school chemistry titration and disease transmission lectures as compared to traditional classroom learning. They found significantly greater learning gains for the MR learning sessions as compared to the small to moderate learning gains in the regular instruction sessions. The authors suggested that the increased levels of embodiment and collaboration were the driving forces behind MR's effectiveness. Here, embodiment is characterized by the necessary integration of motor and sensory systems to learn and perform a task. The authors introduced a system of taxonomy to describe the level of embodied learning provided by VE platforms and proposed three essential components that underlie embodied learning, which contribute to learning effectiveness to different degrees: (1) the amount of engagement of the motor system, (2) the accuracy of mapping gestures to/degree of interaction with the learning content, and (3) the perception of being immersed. Further, studies have shown that 3D platforms allow the participant to (a) manipulate objects or observe objects being manipulated, (b) process action verbs, and (c) observe actions of another individual or avatar, and these aspects are particularly effective in engaging the motor system (Mahon and Caramazza 2008). Corroborating these elements, Schwienhorst (2002) outlines several components of iVR environments that are theoretically effective for L2 learning, including increasing learner autonomy and enabling the L2 learner to organize their learning environment and experience. Here, learner autonomy encompasses the ability to plan and monitor the information being processed and learned in a self-aware manner and includes the ability to move objects around and to interact with various aspects of the learning environment. The author further emphasized that increased levels of interaction in iVR environments is associated with effective, embodied learning.

1.4. The Current Study

On the basis of the discussion so far, we posit that the main consistent aspects of iVR that lend to successful learning are the levels of immersion and interaction of the platforms, and suggest mechanisms by which these features can be attained. In the current study, we designed an iVR environment to enable L2 learning with a high level of immersive and interaction. A high degree of immersion can be achieved through a high field of regard (3D depth, ideally 360-degree panoramic scene) and the use of a head-mounted display. Unlike desktop-based virtual environments, iVR presents a spatial layout that is dynamically changing in accord with the viewer's perspective via a pair of specialized head-mounted displays (HDM or headsets). This enables a participant turning one's head to gain a different view of the same object, as in a real-life environment. Additionally, highly interactive iVR environments can serve to enhance embodied representations. Popular online gaming platforms (e.g., *Second Life*) typically only allow users to interact with the virtual environment through an avatar on desktop or tablet displays, which reduces the embodiment experience of the users. Immersive VR environments that include the use of handsets and hand-held controllers enable

the user to interact with the environment as in real life. See Table 1 for a comparison of the level of immersion and interaction across many VE studies of L2 learning.

Additionally, our study will examine the effects of spatial navigation in L2 iVR training since our iVR environments allow participants to move in a real lab space, aiming to simulate real-life (RL) immersive L2 learning. The most ecologically valid place for individual to learn animal names in a RL situation is a zoo, which naturally involves a high degree of spatial navigation in order for individuals to reach or see all the animals. Our iVR Zoo condition will allow participants to move in real space and teleport along a path to view all the animals. We will compare this condition to items learned in a kitchen iVR setting, which allows for a high degree of interaction through the ability to pick up and move kitchen items, while still offering a moderate level of spatial navigation since participants will also be able to walk around as in a real kitchen (see Methods for details).

So far, no study has tested the specific elements of iVR environments that may lead to L2 learning success while identifying individuals who may optimally benefit from these platforms. The current study aims to address these gaps in the literature by examining learning performance in an iVR platform that requires a higher level of sensorimotor integration, including examining the possible contributions of spatial navigation and manipulation during L2 iVR learning.

The current study employed a mixed counterbalanced design to compare the efficacy of two different L2 learning contexts: iVR learning versus WW paired association learning. We followed the design approach used by Johnson-Glenberg et al. (2014), in which participants learned via both classroom-based and augmented reality-based sessions in a counterbalanced order. Furthermore, we examined the specific components of iVR and WW L2 learning, and the conditions under which each platform was maximally effective. Finally, we examined individual differences in cognitive performance that were associated with or changed as a result of L2 learning context. We predicted that the availability of visuo-spatial features for analysis and physical interaction in iVR environments would provide a distinct learning experience when compared to learning with text or 2D pictures. We further hypothesized that working memory, conflict monitoring ability, and L1 proficiency would be associated with L2 performance.

2. Methods

2.1. Participants

Sixty-four native English monolingual undergraduate students (mean age = 19.05, range = 18–22; 49 female, 15 male) at the Pennsylvania State University participated in this study. All participants provided written consent before any data were collected. Exclusionary criterion included any participants who (1) had any untreated visual or auditory deficits or a history of neurological disorders, (2) were left-handed, as assessed by performance on a Handedness Questionnaire (Snyder and Harris 1993), and (3) had any knowledge of Mandarin Chinese, advanced L2 learning, studied abroad, or had learned a tonal language, as assessed through self-reports on the Language History Questionnaire (LHQ 3.0; Li et al. 2019).

While all participants learned the same 60 words (30 kitchen items and 30 zoo items) across two learning sessions in one iVR session and one WW session, the order and specific learning contexts by which they learned the kitchen and zoo items were counterbalanced. Specifically, Group 1 had 32 participants (mean age = 18.9; 25 female, 7 male) who learned the kitchen items in the iVR context and the zoo items in the WW context, with 16 participants (Group 1A) learning the iVR Kitchen items in the first L2 learning session and WW Zoo items in the second session and 16 participants (Group 1B) learning the WW Zoo items in the first session and iVR Kitchen items in the second session. Group 2 had 32 participants (mean age = 19.2; 24 female, 8 male) who learned the kitchen items in the WW context and the zoo items in the iVR context, with 16 participants (Group 2A) learning the WW Kitchen items first, and 16 learning the iVR Zoo items first (Group 2B; see Figure 1).
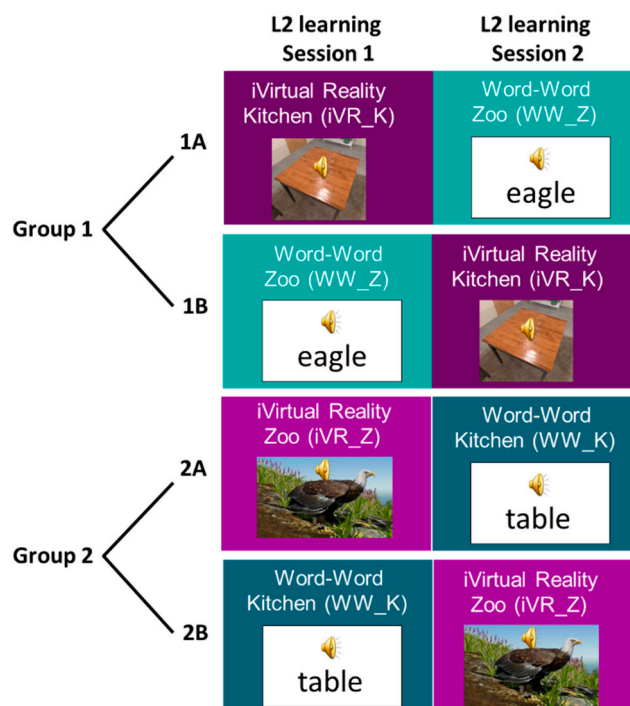
Figure 1. Overview of L2 learning conditions. All participants learned the same 30 kitchen items and 30 zoo items, and all learned 30 words in an immersive virtual reality (iVR) condition and 30 words in a word–word (WW) paired association condition through a counterbalanced design. Participants in Group 1 learned via iVR Kitchen and WW Zoo, with Group 1A learning iVR Kitchen items first and Group 1B learning WW Zoo items first. Participants in Group 2 learned via iVR Zoo and WW Kitchen, with Group 2A learning iVR Zoo items first and Group 2B learning WW Kitchen items first.

2.2. L2 Learning and Testing Materials

All participants learned a total of 60 aurally presented disyllabic Mandarin Chinese words (see Appendix A for full list), with 30 kitchen words comprising items one would typically find in a kitchen (e.g., plate, sink, table) and 30 zoo words comprising common animals one would find in a zoo (e.g., monkey, panda, tiger). These words were selected as a representative sample of what new L2 learners may encounter, were recorded by a native Mandarin Chinese speaker in a sound-attenuated booth, and were used in our previous study comparing two other learning environments (Lan et al. 2014). Across all L2 learning conditions, participants learned the same L2 words and could click on items as many times as they wanted to, within a 20-min time limit.

Stimuli for the WW conditions included English centered text paired with the corresponding Mandarin Chinese word using E-Prime software (Schneider et al. 2002). At the bottom of the screen, participants could see a clock that indicated time remained. Stimuli for the iVR groups included fully immersive environments and items purchased from the Unity store (https://assetstore.unity.com/), TurboSquid (https://www.turbosquid.com/), and CG trader (https://www.cgtrader.com/), which were edited using Unity software. Our VR picture stimuli used during the recognition testing were normed by over 70 participants for picture-naming ability in English, confirming that all pictures can be named with above 90% accuracy (see Appendix B for all 3D pictures used). Using the HTC Vive headgear and handsets, participants could use the right handset to laser point to any animal or object to hear the corresponding Chinese word. Further, participants could move in real space (both iVR environments) or teleport (zoo environment only), and pick up or move objects (kitchen environment only). At the 20, 15, 10, 5, and 1-minute marks, a timer would briefly show at the top of the screen to inform the participants how much time remained.

2.3. Procedure

All participants underwent a total of three sessions for the study, beginning with an hour long cognitive pre-test session on visit 1, followed by the first L2 learning session on visit 2, and ending with a second L2 learning and cognitive post-test session on visit 3. Participants were instructed to learn the Mandarin Chinese words by clicking on objects and animals to hear their corresponding words. All participants performed a practice learning session before each formal L2 learning session to ensure comprehension of the L2 learning task. Across all L2 learning sessions, participants were given 20 min to learn 30 items.

After the practice session and before L2 learning, participants were informed that they would be tested on their recognition accuracy on all 30 words immediately following the L2 learning. This test consisted of a four-alternative forced choice (4AFC) recognition task, where participants would hear a Chinese word and were instructed to click on the English word (for WW conditions) or 3D picture (for iVR conditions) that corresponded to the word they heard. Each word or picture was used as a target once and as a distractor three times in the 4AFC task. Participants received immediate feedback on their performance, via a green rectangle around their selection if they were correct, or a red rectangle around their selection if they were incorrect. Participants also viewed their overall accuracy percentage score at the end of each L2 testing session.

2.3.1. Word–Word (WW) Association Learning

The practice sessions for the WW Kitchen and WW Zoo conditions consisted of four kitchen or zoo practice words, which were not used during the actual L2 learning session. During this practice session, participants were given 1.5 min to learn these four words, and could ask any questions about functionality of the program. Using E-prime software on desktop computers, participants were shown an English word in the middle of the screen simultaneously presented with the corresponding Chinese auditory word. Participants could press the space button to hear the Chinese word again as many times as they chose to, so long as it was within the time limit. Next, they could press the right arrow key to move to the next word. Once they had reached all 30 words, participants in both the WW Kitchen and Zoo conditions could continue to click on the right arrow to keep learning the words in a randomized order as many times as they wanted, again so long as it was within the time limit. Between L2 word presentations, there was a 1–2 second jittered blank screen inter stimulus interval (ISI) for the WW Kitchen condition, and a 2–4 second jittered blank screen ISI for the WW Zoo condition, which functioned to aid in attention and to be more comparable to the iVR learning conditions, where part of the time spent during L2 learning was spent navigating the environments. The actual L2 learning sessions were identical to the practice learning environment except that participants were given 20 min to learn 30 Chinese–English word pairings.

2.3.2. Immersive Virtual Reality (iVR) learning

iVR Kitchen

The practice session for the iVR Kitchen consisted of a virtual kitchen containing four items with red arrows pointing towards them. Participants were instructed to use their right thumb to press on the top button of the HTC Vive handset and aim the laser beam towards one of the items to learn. Once the laser beam would reach an item, participants could hear the corresponding Chinese word for that item (see Figure 2A).
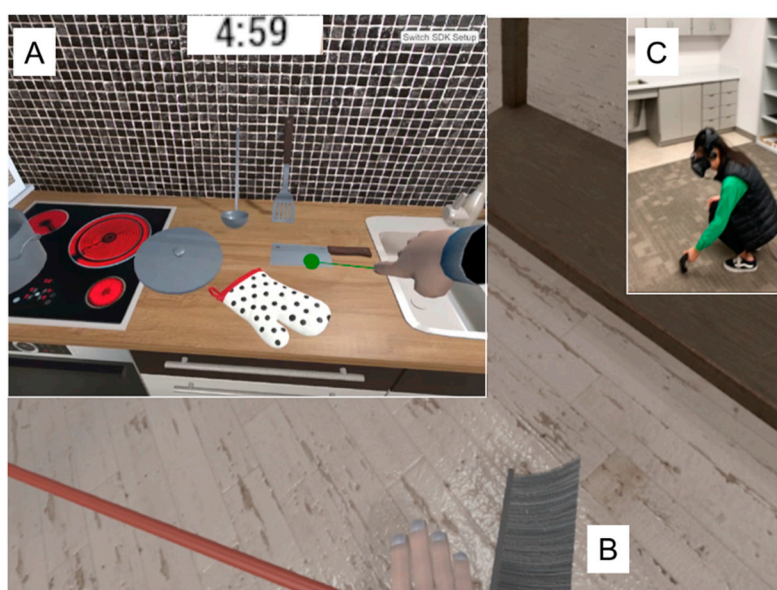
Figure 2. iVR Kitchen interactions. (**A**) Participants used their handset to point to any item and hear their corresponding Chinese word. Across both iVR Kitchen and Zoo conditions, timers briefly appeared every five minutes and during the last five seconds to indicate the time remaining. (**B**) Individuals could pick up and move objects by pressing a trigger button with their index finger. (**C**) Position of an individual picking up the item (broom) shown in panel B.

Participants were then instructed to hear the Chinese words for the remaining items in the practice, where they could click on each of the items as many times as they needed to familiarize themselves with the clicking/laser beam function. Next, participants were instructed to walk over to two of the items and practice picking them up and moving them. This function was enabled by participants moving right next to the item they wanted to manipulate and pressing the lower trigger button with their right index finger to grab on to the object and releasing the trigger button whenever they wanted to set the item back (Figure 2B,C).

Once participants were familiar with both handset functions, they were informed that the actual L2 learning would begin, and that they would be tested on their auditory recognition for the items after 20 min of L2 learning. It was emphasized that the main goal of the L2 learning was to learn the L2 words for each of the 30 items in the kitchen, and that the object manipulation was optional if they found it helped them learn the items. Once the participants were ready to begin the actual L2 learning, they were teleported to the real iVR Kitchen environment where there were 30 kitchen items to learn.

The main differences between the practice kitchen and the actual L2 learning kitchen were that (1) the number of items to learn differed, with four words in the practice and 30 words in the L2 learning, and (2) there were arrows pointing to the four items in the practice whereas there were no arrows at the beginning of L2 learning. However, at the 10- and 15-min marks of learning, arrows would appear next to certain items if participants had not clicked on them yet. This was done to encourage active and self-paced learning at the beginning stage of learning, and then served as a guide to point out any remaining items to be learned during the second half of learning. This procedure ensured that all participants were exposed to all 30 items across learning.

iVR Zoo

The practice session for the iVR Zoo included two parts. The first part was geared towards instruction of the iVR handset functions and the second part functioned to ensure participants were familiar with how to navigate the zoo on an island to reach all 30 animals. The practice and L2 learning environments were identical as in the iVR Kitchen, except that during the practice session, they did not hear any of the Chinese words for any items since the main purpose of the practice sessions were

to habituate the participants to the environment and handset functions. Across the island, hovering gems appeared as indicators of the targets next to any animal to be learned (see Figure 3A).
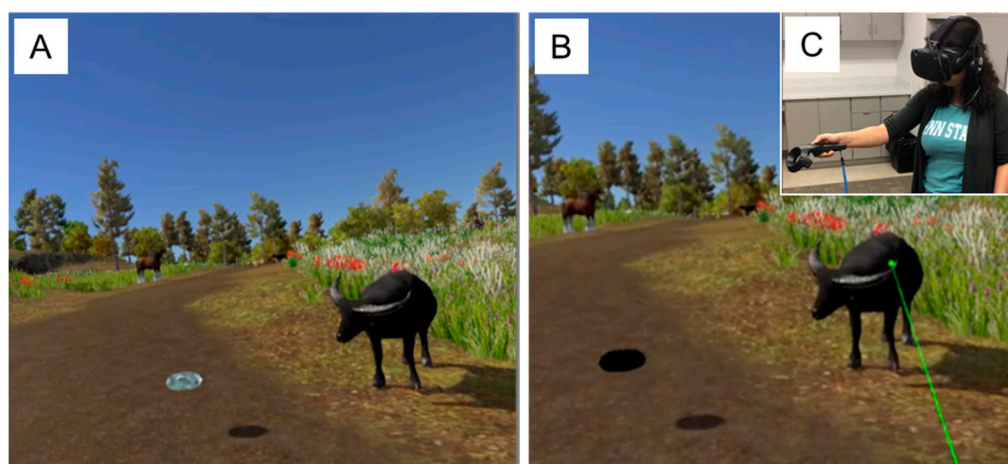

Figure 3. iVR Zoo interactions. (**A**) Participants used their handset to click on animals with a floating gem next to them. (**B**) Once participants successfully clicked on an animal, the gem turned black. During the practice session, no sound was heard when participants clicked on animals. During the actual L2 learning, participants heard the Chinese word for the corresponding animal whenever they clicked on them. (**C**) Position of an individual clicking on the animal in panel B.

During the first part of the practice session, participants could use their right thumb to emit a laser beam which, when it reached an animal, would turn the gem to a darker color. This indicated to the participant that they had successfully clicked on the animal (Figure 3B). Next, participants were shown that they could teleport ahead of themselves to move more efficiently across the island by using their left thumb to press on the top button of the left handset. Once participants pressed this button, a green arc appeared indicating the trajectory of where they would land (see Figure 4B). Participants then navigated across a small section where they could teleport to and click on four animals.
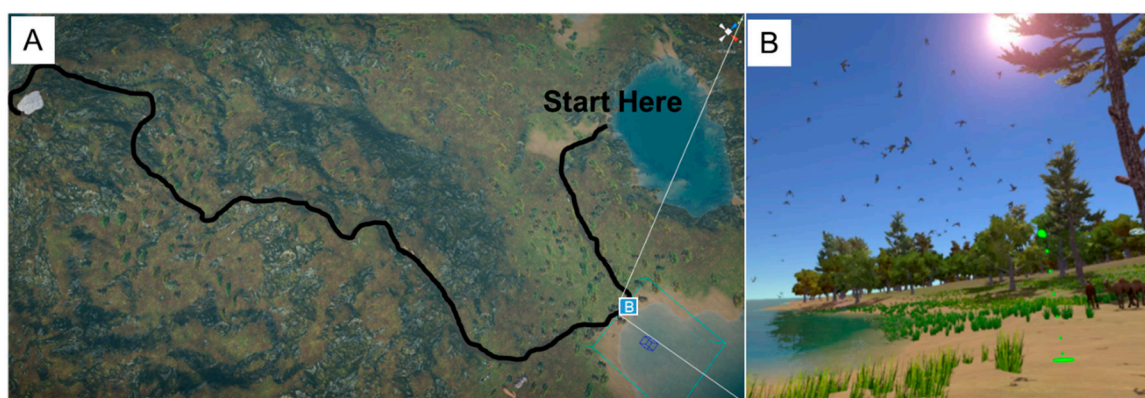

Figure 4. iVR Zoo navigation. (**A**) Participants in the iVR Zoo condition were instructed to follow a path since all animal items to be learned could be seen while navigating this path. (**B**) To navigate more quickly across the island, participants used the teleport function via their left handset: when participants clicked on the top button, they could see a green arc indicating the trajectory of where they would land.

Once participants understood the handset functionalities, they were teleported to the second stage of the practice session, where they followed a winding dirt path which extended across the entire island (see Figure 4A). At this point, participants were instructed to focus on following the path, where they could see all the animals they would be learning during the actual L2 learning session. This was

done to ensure participants were fully proficient with the teleport function and would not get lost during the actual L2 learning. In this fashion, the actual L2 learning time could be largely devoted towards learning the L2 words for the items. The entire island took approximately 3–5 min to traverse, depending on individual preferences. Once participants had reached the top of the island and walked back towards the beginning point (and therefore seen all animals twice), the learning session was complete. This learning session lasted approximately 7–10 min across participants. Participants were then informed that the actual L2 learning would begin and would be followed immediately by a test of auditory recognition of all 30 animals.

During the actual L2 learning session, participants would teleport using their left hand and would click on animals to hear their corresponding Chinese word with their right hand. Participants were informed that they could navigate up and down the island and click on the animals as many times as they chose so long as it was within the 20-min limit. The average navigation time between animals was approximately five seconds, such that participants usually came across a different animal after five seconds of navigating.

The main differences between the iVR Kitchen and iVR Zoo were: (1) the types of items learned, such that the kitchen involved items that could mostly be construed as tools (static), whereas the zoo contained animals (dynamic); (2) the degree of interaction, such that the iVR Kitchen had more interactive options (e.g., clicking on items and moving them) as compared to the iVR Zoo; and (3) the degree of spatial navigation needed to perform the task, such that the iVR group was instructed to follow a winding path, and were able to walk in real space and teleport ahead, whereas the kitchen group only needed to walk in real space since the kitchen items were smaller and therefore closer together (see Figure 5 for an overview).
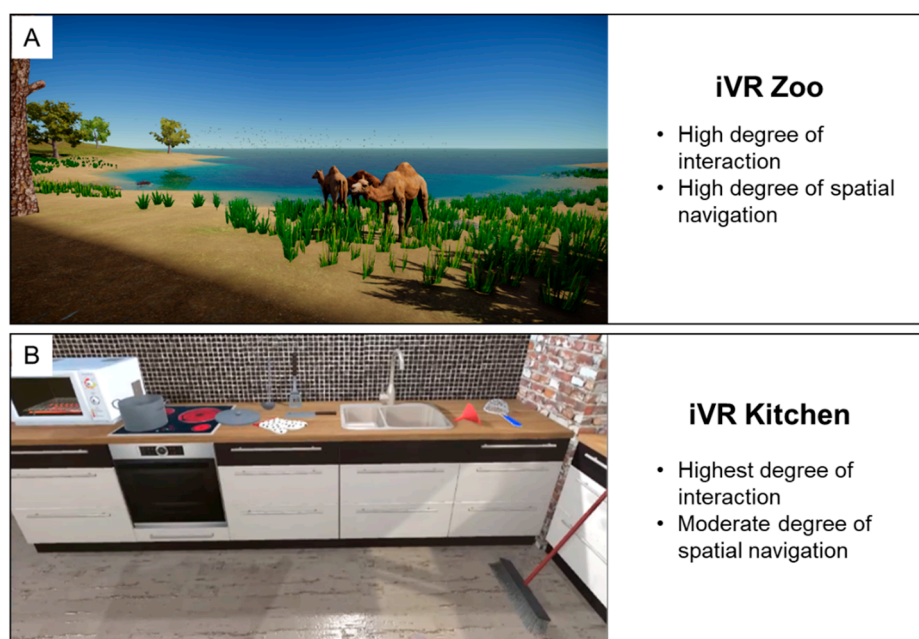

Figure 5. Overview of iVR Kitchen and Zoo differences. (**A**) The iVR Zoo provided a high degree of interaction via clicks on animals to hear their Chinese words and a high degree of spatial navigation through the ability to both walk in real space and use the teleport function. (**B**) The iVR Kitchen provided the highest degree of interaction via clicks on items to hear their Chinese words as well as the ability to move objects, while offering a moderate degree of spatial navigation via the ability to walk in a real space.

2.4. Cognitive Measures

Previous work indicated that L2 learning in VEs could be related to the learner's cognitive and linguistic abilities, including cognitive control and spatial abilities. Specifically, L2 learning ability has been shown to be associated with greater conflict monitoring ability, working memory performance, and L1 proficiency (Abutalebi et al. 2012; Brecht et al. 1995; Linck et al. 2009). We therefore examine whether performance during these tasks is correlated with L2 learning accuracy across our L2 learning conditions.

2.4.1. Attentional Network/Flanker Task (ANT)

This task largely consists of a flanker task, which was used as a measure of inhibitory control and conflict monitoring ability (Fan et al. 2002). Participants were presented with randomized trials containing five arrows or lines on the screen and were asked to indicate the location of the middle (3rd) arrow. This task contained 96 trials comprised equally of three trial types: (1) congruent trials where all arrows pointed towards the same direction, (2) incongruent trials containing a middle arrow pointing in the opposite direction of the surrounding arrows, and (3) neutral trials with a middle arrow pointing in any direction, which was surrounded by arrow-less lines. The main variable of interest for the current study was the traditional flanker effect, therefore we only compared the average reaction time for the incongruent trials versus the congruent trials in our analyses. Participants performed this task before and after L2 learning.

2.4.2. Language History Questionnaire (LHQ)

The LHQ 3.0 (Li et al. 2019) is an updated comprehensive survey on language history and experience. This online tool assesses the language history background through a web-based interface for any and all used and/or learned languages across participants. Language background information included elements such as speaking, writing, comprehension, and usage habits in settings such as at home, in school, and with friends and across all learned languages. Participants completed the LHQ before L2 learning.

2.4.3. Letter Number Sequencing (LNS) Task

This task was adapted from the Wechsler Adult Intelligence Scale (WAIS) and was used as a measure of phonological working memory (Wechsler 1997). During this task, participants heard an auditory sequence of letters and numbers (e.g., k3b9) and were instructed to re-order and type these stimuli in ascending numeric and alphabetic order (e.g., 39bk) on the following screen. Task difficulty gradually increased with each trial, starting with string sequences of two characters and ending with string sequences containing eight characters. There were a total of 21 trials with no time limit for the task, and the measure used in our analyses was the sequence accuracy across trials. All participants performed this task before and after L2 learning.

2.4.4. Peabody Picture Vocabulary Test 4 (PPVT-4)

This task examines English vocabulary aptitude (Dunn and Dunn 2007). Since English was the native language (L1) for all our participants, this test was used as a measure of L1 proficiency. Participants heard an English word which was simultaneously presented with four black and white line drawings. To hear the word again, participants could press the space bar. Beneath each picture was a number (1–4) that corresponded to that picture. Participants were instructed to select the number that best corresponded to the heard word. If they did not know the word, they could press the "5" key. Task difficulty depended on the success of the previous trial. Participants performed this task only before L2 learning.

2.4.5. Spatial Reasoning Instrument (SRI)

This instrument measures mental rotation, spatial orientation, and spatial visualization abilities (Ramful et al. 2017). The multiple-choice questionnaire includes 30 questions in ascending difficulty. Since the task is time-consuming, to reduce practice effects we split this task into two versions. The first version was given before L2 learning and included 15 questions, where five of these questions were each on mental rotation, spatial orientation, and spatial visualization. These questions were easy to medium difficulty questions. The second version was given after L2 learning, where each question was one level of difficulty higher than those included in the first version and included the same number of rotation, orientation, and visualization questions. The only performance measure used was accuracy across trials at pre- and post-test.

2.5. Data Analyses

A binomial logistic regression was used to examine main effects of learning context and learning category using SPSS Statistics (version 24; IBM Corporation: Armonk, NY, USA). This was followed by a second binomial logistic regression to examine individual differences in L1 and cognitive performance on L2 accuracy across trials. Specifically, learning context, learning category, total exposures (clicks on items), L1 performance, working memory improvement, conflict effect, and the interaction between exposures and learning context were all modeled as fixed effects on the target of L2 learning accuracy across trials. Further, subjects, items, age, and gender were modeled as random effects using the intercepts. All categorical variables (learning context, category, and gender) were dummy-coded as zeroes (WW learning context, zoo category, and females) or ones (VR learning context, kitchen category, and males), L2 accuracy was coded as zeroes for incorrect trials and ones for correct trials, and all other variables were included as continuous variables. To examine the learning context effects that may be dependent upon L2 proficiency, we separated the 64 participants into 30 successful learners, who performed with a mean accuracy above 80% across both L2 learning sessions, versus 34 less successful learners, performing with a mean accuracy below 80% across both L2 learning sessions. This split was decided since it was the most accurate method for ensuring there were a comparable number of participants in each comparison group[1]. We then ran exploratory post-hoc tests using the binomial logistic regression as mentioned above separately for the successful learners and the less successful learners. Across all these analyses, we used the family-wise error rate (FWER) to correct for multiple comparisons.

3. Results

3.1. Effects of L2 Learning Context across All Participants

Table 2 presents an overview of the means and standard deviations across learning conditions. Table 3 displays the main effects for the experiment.

Table 2. Overview of means and standard deviations (SD) for L2 accuracy, reaction time (RT), and number of exposures for each L2 learning condition. Average L2 accuracy scores denote the mean percentage of correct responses per condition.

Learning Condition	L2 Accuracy Mean (SD)	L2 Correct RT Mean (SD)	Number of Exposures Mean (SD)
iVR Kitchen	81.15 (11.87)	3638.67 (671.70)	505.69 (118.61)
iVR Zoo	74.38 (19.09)	4061.82 (720.95)	269.31 (87.19)
WW Kitchen	75.10 (20.59)	3734.41 (644.75)	329.03 (137.08)
WW Zoo	68.23 (24.97)	3948.93 (1323.47)	292 (120.79)

---

[1]   If the split was based solely on the mean accuracy (which was 74.7), there would have been 36 participants in the successful learner group and 28 in the less successful learner group, which would not have been as comparable.

Table 3. Overview of main effects from binomial logistic regression across all participants.

	Fixed Effects and Coefficients						
Source	**F**	**df1**	**df2**	**Model Coefficient**	**Standard Error**	**t**	**Sig. ***
Corrected Model	23.634	2	3836				0.000
Intercept				0.864	0.1181	0.550	0.000
Learning context	19.869	1	3836	0.358	0.0951	4.827	0.000 *
Item category	28.331	1	3836	0.428	0.0951	5.021	0.000 *

Abbreviations: df: degrees of freedom; Sig: significance. This applies to this and all remaining tables. * Remains significant after family-wise error rate (FWER) correction. Learning context and item category were dummy-coded as zeroes (WW context, zoo items) and ones (VR context, kitchen items).

There was a significant effect of L2 learning group in L2 learning accuracy (Figure 6A) such that the accuracy for the iVR trials was significantly higher as compared to the WW trials. There was also a significant effect of learning category, such that items learned in the kitchen session were more accurately recognized as compared to items in the zoo session (Figure 6B).
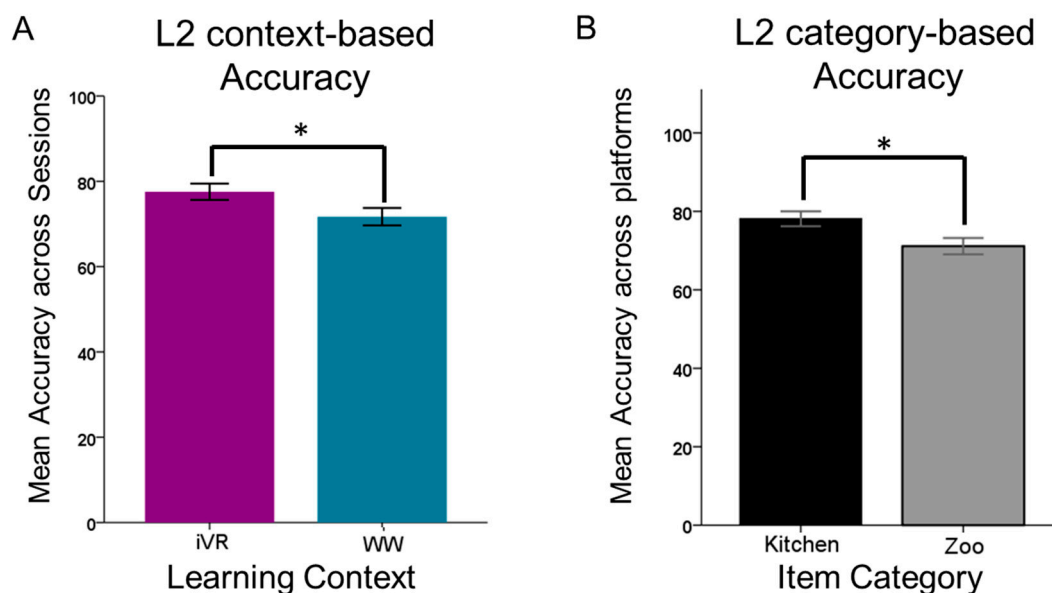
Figure 6. (**A**) There was a main effect of learning context across L2 learning such that accuracy for iVR trials was greater than accuracy for WW trials. (**B**) There was a main effect of item category such that kitchen items were recognized more accurately than zoo items. Error bars indicate 95% confidence intervals (CIs). * indicates significant effect.

3.2. Cognitive Performance Associations with L2 Proficiency

Table 4 presents a summary of means and standard deviations across all cognitive measures. Table 5 presents the individual difference results from the binomial logistic regression. The only significant association between L1 or cognitive ability and L2 performance was that PPVT scores (a measure of L1 proficiency) was significantly positively correlated with L2 accuracy scores for all L2 learners. Higher scores on the PPVT, LNS, and SRI tests all denote higher performance. Further, there was an interaction effect such that the relationship between L2 exposures and L2 accuracy varied by L2 learning context (see Figure 7). ANT scores reflect the conflict effect; therefore lower scores denote higher conflict monitoring and inhibitory control ability.

Table 4. Overview of cognitive performance. Overview of participant population (*n*) and means and standard deviations (SD) for the English Peabody Picture Vocabulary Test (PPVT), attentional network task conflict effect (ANT CE) at pre-test (T1) and post-L2 training test (T2), letter number sequencing (LNS) working memory L2, and overall spatial reasoning instrument (SRI) performance as well as performance on spatial orientation (SRI SO) sub-questions across immersive virtual reality (iVR) and word-word association (WW) learning groups.

Cognitive Measure	Group 1: iVR Kitchen and WW Zoo Mean (SD)	Group 2: iVR Zoo and WW Kitchen Mean (SD)
n	32	32
Gender	25 female 7 male	24 female 8 male
Age	18.9 (0.78)	19.2 (0.90)
PPVT	190.44 (12.47)	192.97 (10.64)
ANT CE T1	119.65 (70.33)	111.89 (66.50)
ANT CE T2	94.95 (47.20)	85.14 (48.55)
LNS T1	0.52 (0.14)	0.55 (0.15)
LNS T2	0.58 (0.13)	0.62 (0.13)
SRI Total T1	0.68 (0.18)	0.68 (0.12)
SRI Total T2	0.65 (0.18)	0.68 (0.15)
SRI SO T1	0.90 (0.14)	0.91 (0.12)
SRI SO T2	0.79 (0.19)	0.91 (0.12)

Table 5. Overview of individual difference results from binomial logistic regression across all participants.

Fixed Effects and Coefficients							
Source	F	df1	df2	Model Coefficient	Standard Error	t	Sig. *
Corrected Model	6.191	9	3641				0.000
Intercept				−7.335	2.2815	−3.215	0.001
LNS 1	0.131	1	3641	0.378	0.363	0.717	0.717
LNS 2	0.002	1	3641	−0.046	−0.042	0.967	0.967
PPVT	9.068	1	3641	0.039	3.011	0.003	0.003 *
SRI_S1	0.024	1	3641	0.161	0.155	0.877	0.877
SRI_S2	0.334	1	3641	0.663	0.578	0.563	0.563
Conflict effect 1	0.293	1	3641	−0.001	−0.542	0.588	0.588
Conflict effect 2	0.634	1	3641	0.003	0.796	0.426	0.426
Learning exposures	10.655	1	3641	0.011	1.160	0.246	0.246
Learning exposures x L2 context	19.714	1	3641	0.030	4.440	0.000	0.000 *

* Remains significant after FWER correction. All fixed coefficients were included as continuous values, except for L2 context which was dummy-coded as zeroes (WW) and ones (VR).

To examine the interaction between L2 context, L2 accuracy, and L2 exposures, we graphed the relationship between average accuracy across trials and total exposures across trials for each L2 learning context (Figure 7). This figure suggests a possible negative correlation between iVR Zoo L2 accuracy and exposures and a possible positive correlation between WW Zoo L2 accuracy and exposures. WW Kitchen and iVR Kitchen contexts show no correlation between L2 accuracy and exposures. The number of L2 exposures was not associated with L2 performance in any learning condition.
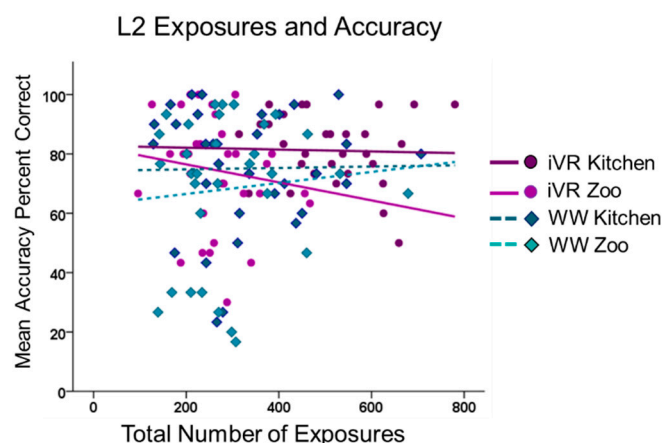
Figure 7. The relationship between L2 exposures and L2 accuracy varied by L2 learning context.

3.3. L2 Context Effects for Successful and Less Successful L2 Learners

Participants were separated into less successful learners (*n* = 34; average accuracy across sessions <80%) and successful learners (*n* = 30; average accuracy across sessions >80%). Table 6 displays the main effects for the high L2 accuracy group. There were no significant effects for this group. Table 7 displays the main effects for the low L2 accuracy group.

Table 6. Overview of main effects from binomial logistic regression in high-accuracy L2 learners.

Fixed Effects and Coefficients for High-Accuracy L2 Learners							
Source	**F**	**df1**	**df2**	**Model Coefficient**	**Standard Error**	**t**	**Sig. ***
Corrected Model	1.433	2	1797	1.938	0.1500	12.921	0.239
Intercept				1.938	0.1500	12.921	0.000
Learning context	0.183	1	1797	0.064	0.1503	0.428	0.669
Item category	2.779	1	1797	0.251	0.1506	1.667	0.096

* Remains significant after FWER correction. Learning context and item category were dummy-coded as zeroes (WW context, zoo items) and ones (VR context, kitchen items).

Table 7. Overview of main effects from binomial logistic regression in low-accuracy L2 learners.

Fixed Effects and Coefficients for Low-Accuracy L2 Learners							
Source	**F**	**df1**	**df2**	**Model Coefficient**	**Standard Error**	**t**	**Sig. ***
Corrected Model	24.756	2	2036				0.000
Intercept				0.065	0.1181	0.550	0.582
Learning context	25.214	1	2036	0.459	0.0951	4.827	0.000 *
Item category	23.304	1	2036	0.478	0.0951	5.021	0.000 *

* Remains significant after FWER correction. Learning context and item category were dummy-coded as zeroes (WW context, zoo items) and ones (VR context, kitchen items).

While there were no significant learning context effects for the successful learners (Figure 8A), this effect was significant for the less successful learners, indicating that participants scored with significantly higher accuracy during iVR sessions (M = 67.5; SD = 14.17) as compared to WW sessions (M = 56.86; SD = 21.51; Figure 8B).
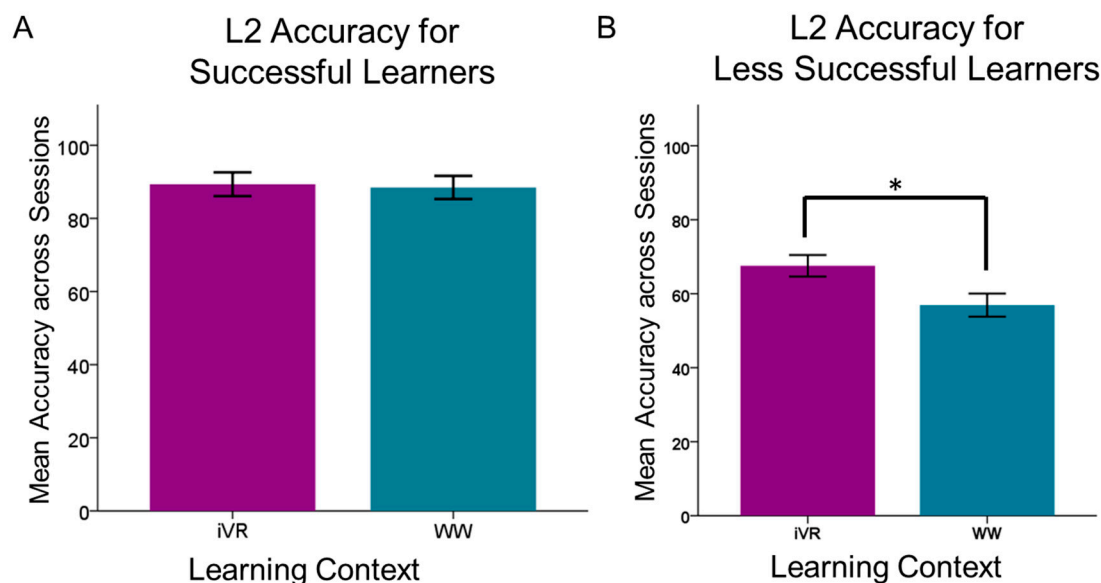
Figure 8. Effects of learning context for successful vs. less successful L2 Learners. (**A**) There were no significant learning environment effects for more successful learners. (**B**) Less successful L2 learners performed with higher accuracy in iVR learning environments as compared to WW learning environments. Error bars indicate 95% CI. * indicates significant effect.

Next, we examined whether these cognitive measures were associated with L2 accuracy in our low-accuracy participants alone, across all L2 learning conditions. Reflecting our results in all participants, only PPVT and the interaction between L2 exposures and L2 context were associated with L2 learning accuracy (see Table 8).

Table 8. Overview of individual difference results from binomial logistic regression in low-accuracy L2 learners.

Source	F	df1	df2	Model Coefficient	Standard Error	t	Sig. *
Fixed Effects and Coefficients for Low-Accuracy L2 Learners							
Corrected Model	6.191	9	3641				0.000
Intercept				−4.631	1.7089	−2.710	0.007
LNS 1	0.617	1	1904	0.804	1.0234	0.786	0.432
LNS 2	0.163	1	1904	−0.345	0.8547	−0.404	0.686
PPVT	5.280	1	1904	0.022	0.0095	2.298	0.022
SRI_S1	0.029	1	1904	0.133	0.7873	0.169	0.866
SRI_S2	0.024	1	1904	0.137	0.8779	0.155	0.876
Conflict effect 1	0.072	1	1904	0.000	0.0018	−0.269	0.788
Conflict effect 2	0.935	1	1904	0.002	0.0023	0.967	0.334
Learning_exposures	14.551	1	1904	0.019	0.0110	1.714	0.087
Learning exposures × L2 context	17.472	1	1904	0.033	0.0079	4.180	0.000

* Remains significant after FWER correction. All fixed coefficients were included as continuous values, except for L2 context which was dummy-coded as zeroes (WW) and ones (VR).

4. Discussion

This study set out to identify the effects of iVR for second language vocabulary learning, as compared with traditional classroom-based methods such as word–word paired associations between L2 and L1 words. Our findings suggest that overall, L2 learners show an additional benefit of iVR learning over WW paired association learning, attesting the utility of iVR as a useful context for technology-based instruction in the case of L2 instruction.

4.1. Virtual Reality Platforms Optimally Benefit Less Successful L2 Learners

Our findings indicate a main effect of L2 learning context, where we found that across all participants, accuracy for items learned via iVR was significantly higher as compared to items learned via the WW condition. This finding supports the embodied cognition hypothesis (Barsalou 2008) that learning in contexts involving a high degree of perceptual and sensorimotor integration is more effectively, and is consistent with evidence from other VE studies indicating the effectiveness of these platforms for L2 learning (Jarmon et al. 2009; Lan et al. 2014; Ibáñez et al. 2011; Johnson-Glenberg et al. 2014). Our findings are also in line with studies of enriched environments and their long lasting and wide-ranging benefits in the mind and brain (Lee and Wong 2008; Lövdén et al. 2013). For example, in a previous neuroimaging study, we found that L2 learning performance in the VE group was associated with cortical thickness in the right IPL. This region is implicated in embodied cognition networks and has been shown to be associated with integrating sounds and actions (McNamara et al. 2008). Our main effect findings that trials learned in the iVR environment were more accurately recognized than items in the WW condition further emphasize that using virtual environments maybe be particularly effective, possibly due to the use of more embodied brain networks during L2 learning.

Since the main goals of the current study were to examine the conditions under which iVR L2 learning would be maximally effective, our study further examined learning performance separately for successful versus less successful learners. Our findings showed that the main effect of learning context may be driven by the performance in our low-accuracy learners, who showed a clear benefit of iVR training over WW training. On the other hand, the highest L2 performers in our study were individuals who performed equally well in the iVR and WW sessions, perhaps reflecting a natural predisposition and aptitude towards L2 learning in these individuals. This finding is particularly interesting since most RL immersive contexts require a stronger base foundation in the L2 for people to benefit from the immersion (Brecht et al. 1995). That is, RL immersion more strongly benefits individuals who are already higher performing L2 learners. In contrast, our data suggest that iVR environments may uniquely be able to help even novice-level individuals. It is important to note that the current study only examines short-term L2 vocabulary acquisition and may not generalize towards the entire scope of L2 learning; more research is needed to examine other components of L2 learning, such as grammar and morphology to see if iVR environments can enhance learning across various aspects of L2 performance.

The factors of iVR that lead to effective L2 learning may include the high degree of interaction and immersion since these were the key aspects of RL immersion we pioneered our iVRs to closely resemble. Kitchen items were more accurately learned as compared to zoo items overall. There could be several interpretations to this finding, including the theory that items that are more manipulable, such as tools (as are the majority of kitchen items), lend to greater embodied representations (Martin et al. 1996), which may be particularly effective for L2 learning. In particular, Martin et al. (1996) examined activation of neural networks in response to 2D drawings of tools, animals, and nonsense objects. The nonsense objects were drawings of items that combined features of regular objects to form a novel object and formed a baseline condition for comparison of tools and animals. They found that tools uniquely recruited motor regions of the brain, including brain regions implicated in arm movements associated with tool usage along with language regions associated with generation of action words. On the other hand, animal words recruited additional visual processing regions. Taken together, items that can be categorized as tools may involve more distributed embodied networks, and may be more effectively learned as compared to animal words.

We examined whether the level of exposures was associated with L2 accuracy in all the learning contexts and did not find any significant results. Instead, the relationship between L2 accuracy and L2 exposures may be mediated by the L2 learning context, since the interaction between these three variables was significant. Figure 7 shows the relationship between mean accuracy across trials and total exposures for each group, indicating a possible positive correlation between WW zoo exposures

and L2 accuracy. On the other hand, the iVR Zoo group shows a possible negative correlation between exposures and L2 performance, indicating these participants needed fewer exposures to L2 items to perform well. It is interesting to note that in Lan et al. (2014), amount of exposure also did not correlate with learning success, but the virtual learners only had half of the exposure compared with the non-virtual learners. Hsiao et al. (2017) examined which objects L2 learners clicked in succession when learning via VE. They separated their participants into high-achieving learners (those performing above 85% accuracy at post-test) and low-achieving learners (performing below 85% at post-test) to examine differences in clicking strategy across these learners. Their analyses indicated that more successful learners were those participants who used a "cluster strategy", where the participants' learning sequence was based on object features or sound features (i.e., clicking on objects with L2 auditory stimuli that sound alike in a row; Hsiao et al. 2017), as compared with the less successful learners who tended to use a "nearest strategy" (i.e., clicking on the objects nearby linearly). In the current study, since not all our participants used the manipulation function in the kitchen environment, there were not enough trials to examine the effect of manipulation across our iVR participants. Future studies along these aspects are therefore needed.

4.2. Relationships between L2 Success and Individual Difference Measures

While we expected to find cognitive factors such as working memory capacity and conflict resolution to be associated with L2 performance, none of these relationships was significant. However, across learning platforms, the strongest association with L2 learning success were scores on the PPVT, a test of L1 vocabulary skill. This is consistent with various other studies finding that L1 proficiency is linked with L2 proficiency (Cummins 1991). Further, this result may indicate that the link between L1 and L2 proficiency is not dependent on learning context. Of particular interest was the examination into which factors may underlie the L2 performance for the less successful learners. While we initially predicted that the learning gains in these low performers might reflect greater attentional conflict monitoring ability or working memory performance, as previous studies have found (Abutalebi et al. 2012; Linck et al. 2014; Miyake and Friedman 1998), our findings did not support this hypothesis. This may be due to the high degree of variations participants exhibited during L2 learning, which is difficult to model especially since they only learned via VR and WW for one session each. In a previous study (Legault et al. 2018), we found that working memory predicted performance in VE and not PW learners; however, participants there learned in only one learning context (not both), and had seven sessions to learn the vocabulary. While speculative, additional training sessions or only training in one environment may be important to elicit these relationships.

4.3. iVR Features in Learning that Affect L2 Learning and Cognitive Performance

iVR platforms are uniquely suited to address research questions concerning the specific components of RL immersive learning that foster learning success. According to the taxonomy system proposed by Johnson-Glenberg et al. (2014), our iVR L2 learning platforms would be categorized as enabling the highest level of embodiment. This level of embodiment is characterized by a high degree of locomotion and, a high degree of physical interaction with the objects in the virtual immersive environment. Our L2 context-based results served as a confirmation that iVR environments that have a high level of immersion can be effective for L2 learning. Meanwhile, our item category-based results enabled us to elucidate specific properties of iVR contexts that led to effective L2 learning. During post-test feedback, we asked all participants to indicate which aspects of the learning environment led to effective (or ineffective) L2 learning. Nearly all participants, especially those in the iVR Kitchen condition, stated that the level of engagement was much higher for the iVR contexts as compared to the WW condition. Many (though not all) of the iVR Kitchen condition participants also stated that the ability to move the objects aided their L2 learning process. When asked if they had a preference of learning contexts, the vast majority of participants preferred the iVR conditions; the few who preferred the WW context stated that the reasons were due to complexity: they found the iVR conditions

(especially the iVR Zoo) to be too complex or distracting, while the WW platform was more familiar as an L2-mediated instruction type and was easier.

In conclusion, our study provides a first systematic examination of the effectiveness of using iVR in L2 learning situations. Rather than a simple answer to the efficacy question, our data point to the differential role of iVR for different learners. A possible limitation for our study is that we did not include a picture-word (PW) association condition as a comparison group, although we have already previously compared PW to VE learning conditions (Lan et al. 2014). In future studies, we hope to include more systematic comparisons of VR, VE, PW, and WW learning conditions to further understand the complex relationship between learning context and L2 performance. Together, our findings suggest that highly immersive and interactive virtual environments may enhance embodied experiences, and hence effective learning of L2 materials.

Appendix A

List of all Kitchen and Zoo stimuli used during L2 learning for all participants. Listed here are the English words followed by the Chinese characters (not shown to participants) and the pinyin pronunciation of the words.

Kitchen stimuli:

Bowl	飯碗	(fanwan)
Broom	掃把	(saoba)
Candle	蠟燭	(lazhu)
Chair	椅子	(yizi)
Chopsticks	筷子	(kuàizi)
Clock	時鐘	(shizhong)
Drainer	濾網	(luwang)
Drawer	抽屜	(chouti)
Funnel	漏斗	(loudou)
Jar	罐子	(guanzi)
Knife	刀子	(daozi)
Ladle	湯勺	(tangshao)
Lid	鍋蓋	(guogai)
Mitten	手套	(shoutao)
Oven	烤箱	(kaoxiang)
Plate	盤子	(panzi)
Refrigerator	冰箱	(bīngxiāng)
Scissor	剪刀	(jiandao)

Sink	水槽	(shuicao)
Spatula	鏟子	(chanzi)
Steel cup	鋼杯	(gangbei)
Stewpot	燉鍋	(dunguo)
Stool	凳子	(dengzi)
Stove	電爐	(dianlu)
Table	餐桌	(canzhuo)
Teacup	茶杯	(chabei)
Teapot	茶壺	(chahu)
Telephone	電話	(dianhua)
Vase	花瓶	(huaping)
Wineglass	酒杯	(jiubei)

Zoo stimuli

|		
Bear	黑熊	(heixiong)
Bird	小鳥	(xiaoniao)
Butterfly	蝴蝶	(hudie)
Camel	駱駝	(luotuo)
Cat	小貓	(xiaomao)
Cow	黃牛	(huangniu)
Crab	螃蟹	(pangxie)
Crocodile	鱷魚	(eyu)
Dog	小狗	(xiaogou)
Eagle	老鷹	(laoying)
Elephant	大象	(daxiang)
Fox	狐狸	(huli)
Frog	青蛙	(qingwa)
Goat	山羊	(shānyáng)
Kangaroo	袋鼠	(daishu)
Lion	獅子	(shizi)
Lizard	蜥蜴	(xiyi)
Monkey	猴子	(houzi)
Ostrich	鴕鳥	(tuoniao)
Panda	熊貓	(xiongmao)
Parrot	鸚鵡	(yingwu)
Peacock	孔雀	(kongque)
Penguin	企鵝	(qie)
Rabbit	兔子	(tuzi)
Rhinoceros	犀牛	(xiniu)
Rooster	公雞	(gongji)
Tiger	老虎	(laohu)
Turkey	火雞	(huoji)
Turtle	烏龜	(wugui)
Zebra	斑馬	(banma)

Appendix B

Pictures of all Kitchen and Zoo stimuli used during L2 learning and testing for iVR conditions.
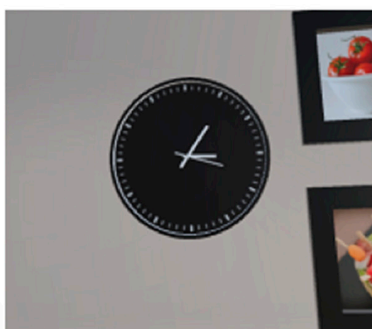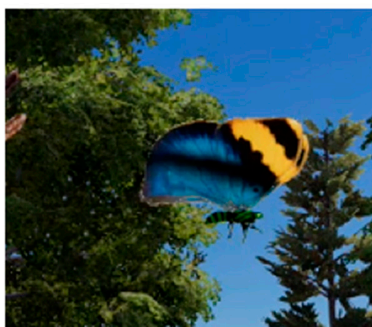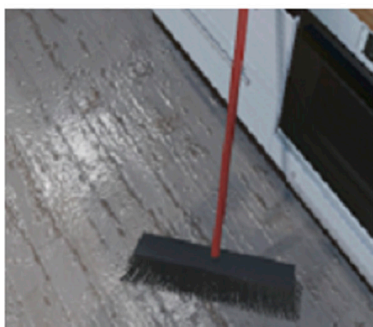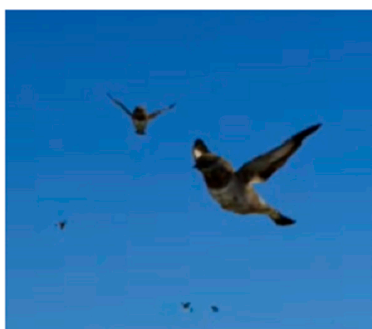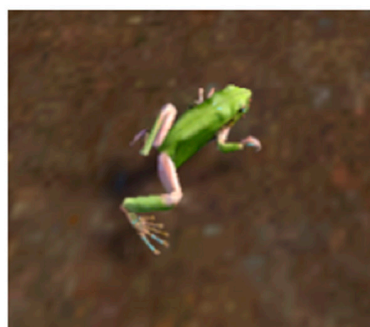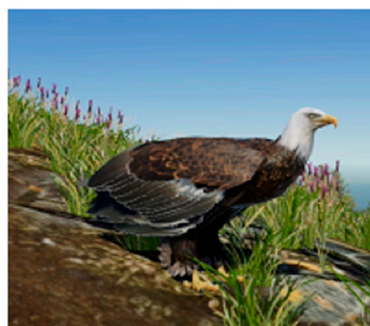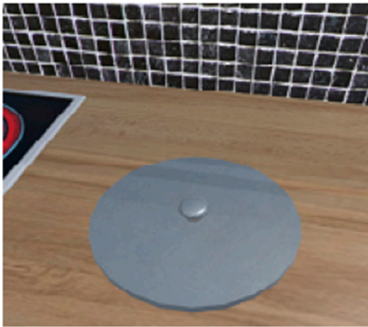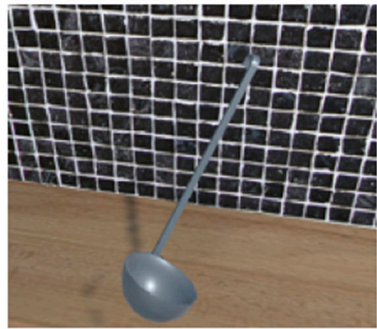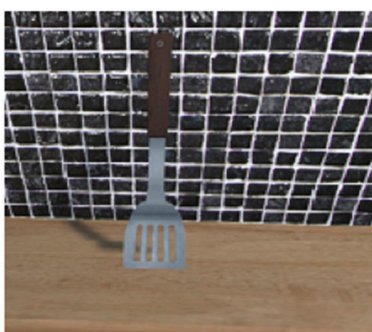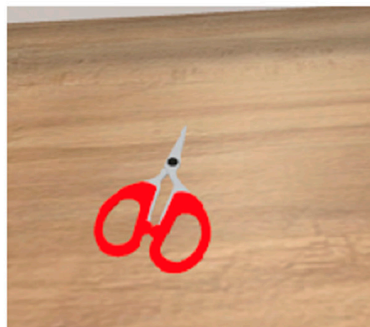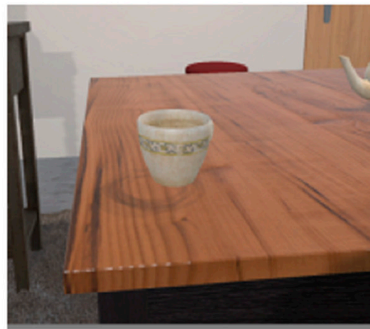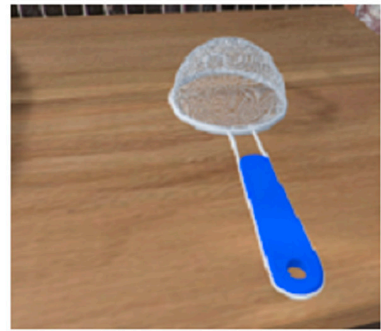
References

Abutalebi, Jubin, and David Green. 2007. Bilingual Language Production: The Neurocognition of Language Representation and Control. *Journal of Neurolinguistics* 20: 242–75. [CrossRef]

Abutalebi, Jubin, Pasquale Anthony Della Rosa, David W. Green, Mireia Hernandez, Paola Scifo, Roland Keim, Stefano F. Cappa, and Albert Costa. 2012. Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring. *Cerebral Cortex* 22: 2076–86. [CrossRef] [PubMed]

Adamovich, Sergei V., Gerard G. Fluet, Eugene Tunik, and Alma S. Merians. 2009. Sensorimotor Training in Virtual Reality: A Review. *NeuroRehabilitation* 25: 29–44. [CrossRef] [PubMed]

Aziz-Zadeh, Lisa, and Antonio Damasio. 2008. Embodied Semantics for Actions: Findings from Functional Brain Imaging. *Journal of Physiology Paris* 102: 35–39. [CrossRef] [PubMed]

Barsalou, Lawrence W. 2008. Grounded Cognition. *Annual Review of Psychology* 59: 617–45. [CrossRef] [PubMed]

Barsalou, Lawrence W., W. Kyle Simmons, Aron K. Barbey, and Christine D. Wilson. 2003. Grounding Conceptual Knowledge in Modality-Specific Systems. *Trends in Cognitive Sciences* 7: 84–91. [CrossRef]

Berns, Anke, Antonio Gonzalez-Pardo, and David Camacho. 2013. Game-like Language Learning in 3-D Virtual Environments. *Computers and Education* 60: 210–20. [CrossRef]

Blascovich, Jim, Jack Loomis, Andrew C. Beall, Kimberly R. Swinth, L. Crystal, Source Psychological Inquiry, and Jim Blascovich. 2002. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry* 13: 103–24. [CrossRef]

Brecht, Richard D., Dan Davidson, and Ralph B. Ginsberg. 1995. Predictors of Foreign Language Gain during Study Abroad. *Second Language Acquisition in a Study Abroad Context* 9: 53–82.

Brooks, Barbara M., Jane E. McNeil, F. David Rose, Richard J. Greenwood, Elizabeth A. Attree, and Antony G. Leadbetter. 1999. Route Learning in a Case of Amnesia: A Preliminary Investigation into the Efficacy of Training in a Virtual Environment. *Neuropsychological Rehabilitation* 9: 63–76. [CrossRef]

Casasanto, Daniel, and Kyle Jasmin. 2018. Virtual Reality. In *Research Methods in Psycholinguistics and the Neurobiology of Language*. Hoboken: John Wiley & Sons, pp. 174–90.

Chen, Julian Cheng Chiang. 2016. The Crossroads of English Language Learners, Task-Based Instruction, and 3D Multi-User Virtual Learning in Second Life. *Computers and Education* 102: 152–71. [CrossRef]

Collentine, J., and Freed B. F. 2004. Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition* 26: 153–171.

Costello, Patrick J. 1997. *Health and Safety Issues Associated with Virtual Reality: A Review of Current Literature*. Loughborough: Advisory Group on Computer Graphics.

Cummins, Jim. 1991. Interdependence of First- and Second-Language Proficiency in Bilingual Children. In *Language Processing in Bilingual Children*. Cambridge: Cambridge University Press, pp. 70–89. [CrossRef]

Dede, Chris. 2009. Immersive Interfaces for Engagement and Learning. *Science* 323: 66–99. [CrossRef] [PubMed]

Dunn, Lloyd M., and Douglas M. Dunn. 2007. *Peabody Picture Vocabulary Test*. Minneapolis: Pearson. [CrossRef]

Fan, Jin, Bruce D. McCandliss, Tobias Sommer, Amir Raz, and Michael I. Posner. 2002. Testing the Efficacy and Independence of Attentional Networks. *Journal of Cognitive Neuroscience* 14: 340–47. [CrossRef] [PubMed]

Fischer, Martin H., and Rolf A. Zwaan. 2008. Embodied Language: A Review of the Role of the Motor System in Language Comprehension. *Quarterly Journal of Experimental Psychology* 61: 825–50. [CrossRef] [PubMed]

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: 275–301. [CrossRef]

Glenberg, Arthur, Marc Sato, Luigi Cattaneo, Lucia Riggio, Daniele Palumbo, and Giovanni Buccino. 2008. Processing Abstract Language Modulates Motor System Activity. *Quarterly Journal of Experimental Psychology* 61: 905–19. [CrossRef] [PubMed]

Hernandez, Arturo, and Ping Li. 2007. Age of Acquisition: Its Neural and Computational Mechanisms. *Psychological Bulletin* 133: 638–50. [CrossRef] [PubMed]

Hernandez, Arturo, Ping Li, and Brian MacWhinney. 2005. The Emergence of Competing Modules in Bilingualism. *Trends in Cognitive Sciences* 9: 220–25. [CrossRef] [PubMed]

Hsiao, Indy Y. T., Yu-Ju Lan, Chia-Ling Kao, and Ping Li. 2017. Visualization Analytics for Second Language Vocabulary Learning in Virtual Worlds. *Journal of Educational Technology & Society* 20: 161–75.

Ibáñez, María Blanca, José Jesús García, Sergio Galán, David Maroto, and Diego Morillo. 2011. Design and Implementation of a 3D Multi-User Virtual World for Language Learning. *Educational Technology & Society* 14: 2–10.

Jarmon, Leslie, Tomoko Traphagan, Michael Mayrath, and Avani Trivedi. 2009. Virtual World Teaching, Experiential Learning, and Assessment: An Interdisciplinary Communication Course in Second Life. *Computers and Education* 53: 169–82. [CrossRef]

Jerald, Jason. 2016. *The VR book: Human-Centered Design for Virtual Reality*. New York: ACM Books, San Rafael: Morgan & Claypool.

Jeong, Hyeonjeong, Motoaki Sugiura, Yuko Sassa, Keisuke Wakusawa, Kaoru Horie, Shigeru Sato, and Ryuta Kawashima. 2010. Learning Second Language Vocabulary: Neural Dissociation of Situation-Based Learning and Text-Based Learning. *NeuroImage* 50: 802–9. [CrossRef] [PubMed]

Johnson-Glenberg, Mina C., David A. Birchfield, Lisa Tolentino, and Tatyana Koziupa. 2014. Collaborative Embodied Learning in Mixed Reality Motion-Capture Environments: Two Science Studies. *Journal of Educational Psychology* 106: 86–104. [CrossRef]

Johnson, Jacqueline S., and Elissa L. Newport. 1989. Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology* 21: 60–99. [CrossRef]

Lafford, Barbara A. 1995. Getting into, through, and out of a Survival Situation: A Comparison of Communicative Strategies Used by Students Studying Spanish-Abroad and 'at Home'. *Second Language Acquisition in a Study Abroad Context* 9: 97–121.

Lan, Yu-Ju, Nian-Shing Chen, Ping Li, and Scott Grant. 2015. Embodied Cognition and Language Learning in Virtual Environments. *Educational Technology Research and Development* 63: 639–44. [CrossRef]

Lan, Yu-Ju, Shin-Yi Fang, Jennifer Legault, and Ping Li. 2014. Second Language Acquisition of Mandarin Chinese Vocabulary: Context of Learning Effects. *Educational Technology Research and Development* 63: 671–90. [CrossRef]

Lee, Elinda Ai-lim, and Kok Wai Wong. 2008. A Review of Using Virtual Reality for Learning. In *Transactions on Edutainment I*. Berlin/Heidelberg: Springer, pp. 231–41. [CrossRef]

Legault, Jennifer, Shin-Yi Fang, Yu-Ju Lan, and Ping Li. 2018. Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition*. [CrossRef] [PubMed]

Lenneberg, Eric H. 1967. *The Biological Foundations of Language*. New York: Wiley. [CrossRef]

Levak, Natasha, and Jeong Bae Son. 2017. Facilitating Second Language Learners' Listening Comprehension with Second Life and Skype. *ReCALL* 29: 200–218. [CrossRef]

Li, Ping, Fan Zhang, Anya Yu, and Xiaowei Zhao. 2019. Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*. in press.

Linck, Jared A., Judith F. Kroll, and Gretchen Sunderman. 2009. Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. *Psychological Science* 20: 12, 1507–1515. [CrossRef] [PubMed]

Linck, Jared A., Peter Osthus, Joel T. Koeth, and Michael F. Bunting. 2014. Working Memory and Second Language Comprehension and Production: A Meta-Analysis. *Psychonomic Bulletin & Review* 21: 861–83. [CrossRef]

Lohse, Keith R., Courtney G.E. Hilderman, Katharine L. Cheung, Sandy Tatla, and H. F. Machiel Van Der Loos. 2014. Virtual Reality Therapy for Adults Post-Stroke: A Systematic Review and Meta-Analysis Exploring Virtual Environments and Commercial Games in Therapy. *PLoS ONE* 9: e93318. [CrossRef] [PubMed]

Lövdén, Martin, Elisabeth Wenger, Johan Mårtensson, Ulman Lindenberger, and Lars Bäckman. 2013. Structural Brain Plasticity in Adult Learning and Development. *Neuroscience and Biobehavioral Reviews* 37: 2296–310. [CrossRef] [PubMed]

MacWhinney, Brian. 2012. The Logic of the Unified Model. In *Handbook of Second Language Acquisition*. London: Routledge, pp. 211–27.

Mahon, Bradford Z., and Alfonso Caramazza. 2008. A Critical Look at the Embodied Cognition Hypothesis and a New Proposal for Grounding Conceptual Content. *Journal of Physiology Paris* 102: 59–70. [CrossRef] [PubMed]

Martin, Alex, Cheri L. Wiggs, Leslie G. Ungerleider, and James V. Haxby. 1996. Neural Correlates of Category-Specific Knowledge. *Nature* 379: 649. [CrossRef] [PubMed]

McNamara, Adam, Giovanni Buccino, Mareike M. Menz, Jan Gläscher, Thomas Wolbers, Annette Baumgärtner, and Ferdinand Binkofski. 2008. Neural dynamics of learning sound—action associations. *PLoS ONE* 3: e3845. [CrossRef] [PubMed]

Milgram, Paul, and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual-Displays. *Ieice Transactions on Information and Systems* 77: 1321–29.

Milton, James, and Paul Meara. 1995. How Periods Abroad Affect Vocabulary Growth in a Foreign Language. *International Journal of Applied Linguistics* 107–108: 17–34. [CrossRef]

Miyake, Akira, and Naomi P Friedman. 1998. Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude. *Foreign Language Learning Psycholinguistic Studies on Training and Retention* 339–64.

Ramful, Ajay, Thomas Lowrie, and Tracy Logan. 2017. Measurement of Spatial Ability: Construction and Validation of the Spatial Reasoning Instrument for Middle School Students. *Journal of Psychoeducational Assessment* 35: 709–27. [CrossRef]

Richardson, Fiona M., and Cathy J. Price. 2009. Structural MRI studies of language function in the undamaged brain. *Brain Structure and Function* 213: 511–23. [CrossRef] [PubMed]

Schubert, Thomas, Frank Friedmann, and Holger Regenbrecht. 1999. Embodied Presence in Virtual Environments. In *Visual Representations and Interpretations*. London: Springer, pp. 269–78. [CrossRef]

Schwienhorst, Klaus. 2002. Why Virtual, Why Environments? Implementing Virtual Reality Concepts in Computer-Assisted Language Learning. *Simulation and Gaming* 33: 196–209. [CrossRef]

Schneider, Walter, Amy Eschman, and Anthony Zuccolotto. 2002. *E-Prime: User's Guide*. Pittsburgh: Psychology Software Incorporated.

Si, Mei. 2015. A Virtual Space for Children to Meet and Practice Chinese. *International Journal of Artificial Intelligence in Education* 25: 271–90. [CrossRef]

Snyder, Peter J., and Lauren Julius Harris. 1993. Handedness, Sex, Familial Sinistrality Effects on Spatial Tasks. *Cortex* 29: 115–34. [CrossRef]

Stein, Maria, Carmen Winkler, Anelis Kaiser, and Thomas Dierks. 2014. Structural brain changes related to bilingualism: Does immersion make a difference? *Frontiers in Psychology* 5: 1116. [CrossRef] [PubMed]

Wang, Yi Fei, Stephen Petrina, and Francis Feng. 2017. VILLAGE—Virtual Immersive Language Learning and Gaming Environment: Immersion and Presence. *British Journal of Educational Technology* 48: 431–50. [CrossRef]

Wechsler, David. 1997. *Wechsler Adult Intelligence Scale III*. San Antonio: The Psychological Corporation. [CrossRef]

Willems, Roel M., and Daniel Casasanto. 2011. Flexibility in Embodied Language Understanding. *Frontiers in Psychology* 2: 1–11. [CrossRef]