Understanding the distinctions in, and impacts of, hand-based sensorimotor interaction in immersive virtual reality; via second language learning

Jack Ratcliffe

Cognitive Science Laboratory
School of Electronic Engineering and Computer Science
Queen Mary University of London

2022

Abstract

It is widely accepted that immersive virtual reality (IVR) depends on head-based sensorimotor interaction, and the implications and impacts of this interaction are well-explored. However, an additional sensorimotor interaction found in many contemporary IVR experiences, hand-based sensorimotor interaction (HBSI), has received far less attention. This is a notable gap in literature, as in the physical world, HBSI is strongly linked with cognition and cognitive outcomes; and is particularly linked with second language learning. This thesis explores HBSI in IVR by examining whether different implementations of HBSI impact cognitive outcomes; and whether cognitive outcomes from HBSI in IVR are congruent with HBSI in the physical world. These findings are also used to comment on how users cognitively perceive the sensorimotor actions they take in virtual environments, as well as on theories of embodied cognition.

Acknowledgements

It is, perhaps, the researcher's greatest burden that they may never *exhaustively* document something. So too is that the case with these acknowledgements; in which it would be exhausting to outline all the contributors to this research and the gratitude I feel towards them. And yet, in true scholarly fashion, I shall attempt to create some facsimile of the actual situation, and note it down as comprehensively as I am able.

To start, I am grateful to my supervisor, Dr Laurissa Tokarchuk, for her guidance and belief in this research and in my development as an academic. Thank you for granting me a field to explore, and for helping fence off the muddiest parts. I would also like to mention my peers in Dr Tokarchuk's research group: Vivek, for his warmth and guidance; Nick, for his passionate rigour; and Gideon and Ollie for their interest and enthusiasm.

I am also grateful to Dr Agnieszka Lyons, for her help in the formation of this work; and for including me in the School of Linguistics; to Prof Pat Healey, for his guidance when needed; and to Prof Nick Brynn-Kinns, for his careful consideration as an independent assessor, his collegiality in our combined research projects, and his leadership of the Media and Arts Technology laboratory.

I could not imagine doing this PhD without the Media and Arts Technology laboratory and its community, a beacon of friendship, enthusiasm and curiosity. From the supportive staff - Jonathan Winfield and Geetha Bommireddy - whose knowledge and efficacy greatly reduced any of the administrative burdens one may have expected to face; to my fellow students, whose humour and friendship I will forever value.

In particular, I offer thanks to six of my peers and friends for exemplifying the traits needed to be an academic: Sophie, for her sense of dedication and perfectionism; Tom, for his earnestness about research; Jorge, for his focus and drive; Mei, for her bravery and enthusiasm; and Teo and Sebastian, for remembering the importance of going outside. I would also like to thank my friend Rei Wada, from the School of Languages, Linguistics and Film, whose kind words and confidence in my work was consistently humbling.

Outside of academia, I have to thank my family, who have offered limitless support to the worst version of me over the last four years. To my mother, Christina, thank you for always believing in me. I am so grateful you are here to see this completed. To my father, Christopher, thank you for our first computer, and the path that it set me down. To my brothers, Caley and Callan, whose influence has shaped me in numerous ways, likely none of which I would admit to; and to Neil, for everything.

Finally, I would like to thank Marc, my happiest distraction.

Licence

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research motivation

The preceding decades of immersive virtual reality (IVR) research have explored a myriad of factors regarding IVR experiences, and the impact of those factors on users. Perhaps the most explored factor is head-based sensorimotor interaction, in which the users' head orientation (and commonly, position) are detected and applied to a virtual camera, allowing users to somewhat naturalistically look around a virtual environment by moving their head.

Head-based sensorimotor interaction is the pre-eminent experiential factor in IVR, and the key driver of the well-studied IVR-related cognitive phenomena, presence (see Chapter 2.3 for a discussion of definitions of IVR-related presence).

Head-based sensorimotor interaction, however, is not the only sensorimotor interaction available when engaging with IVR. Hand-based sensorimotor interaction (HBSI), enabled by tracking the position and orientation of the users' hands (as well as various abstracted forms of interaction, such as pushing a button to simulate 'grabbing'), has become an important form of interaction, particularly for interaction-oriented IVR.

Some researchers believe that HBSI could play an important role in both how users experience IVR, and the outcomes from those experiences. Steuer noted that the "mapping" offered by HBSI and other human actions within a "gloves'n'goggles" mediated environment influences interactivity, and that the more interactive an environment is, the greater the sense of presence evoked by it [309].

However, contemporary views suggest HBSI is not just a driver of presence, but a unique contributor to the IVR experience and outcomes. An increasingly popular theory grounding perspectives on HBSI in IVR is that HBSI plays a special role in human interaction [100]; and that the embodied affordances of gesture and manipulation are a unique profound affordance of IVR that is distinct from presence, and ones which offers strong potential learning benefits [141], particularly when compared with other types of IVR interaction, such as gaze-interaction. A similar perspective is presented in Makransky & Petersen's Cognitive Affective Model of Immersive Learning, in which IVR control factors (of which HBSI is one aspect) impact both presence and agency, which then influence interest, motivation, self-efficacy, embodiment, cognitive load and self-regulation, all of which impact various learning outcomes from IVR experiences [203].

Compared with head-based sensorimotor interaction, the specific cognitive impacts of HBSI on users are under-explored, and it is not yet clear which experiential factors and user outcomes could be affected by HBSI. This is because many of the existing comparative studies of IVR have not isolated head-based sensorimotor interactivity and HBSI [204], often comparing an IVR system with both head-based and hand-based sensorimotor interaction with a control with neither.

One user outcome that may be particularly sensitive to HBSI in IVR, and therefore could indicate whether HBSI could be a potential effector of user experiences in IVR, is learning. There is strong theoretical and experimental support for a relationship between enhanced forms of embodiment, embodied interaction, and learning outcomes [345]. Researchers exploring the potential impact of HSBI on learning have suggested that it could be the the "second profound affordance of immersive virtual reality" (after head-based sensorimotor interaction) [141]; as well as impacting embodied learning, learning, feelings of agency; physical presence; intrinsic motivation; self-efficacy; extraneous cognitive load interaction; extraneous cognitive load environment and situational interest [242]. See Figure 1.1 for an illustration of the factors that head-based and hand-based sensorimotor interaction affect, and which are of particular interest in this thesis.

Therefore measuring learning outcomes from a HBSI learning system could be a useful way of understanding if HBSI can have an impact on IVR users; especially as learning can be measured objectively through a testing process.

One of the topics of learning that is closely linked with sensorimotor activity is second language learning [193][196][81]. Research in this area is predominantly based on the perspective that language in the brain is represented by a sensorimotor network formed from experiences collected on the concept being memorised [250]. In essence, language cognition is sensorimotor and environmentally embodied. Further evidence for this is presented in investigations that show that concrete words (words representing objects that are more easily physically tangible and manipulable via human sensorimotor systems) are easier to remember than abstract ones [36]. The explanation offered for the difference in memorisation between concrete and non-concrete words being that concrete words are richer in potential sensorimotor stimulus.

There have been many experimental studies that show that learning that leverages HBSI (outside of IVR) leads to better second language memorisation [97][96][94][149][2][193][192]. There have also been studies into types of HBSI leading to differing levels of effect; for example, with actions and gestures providing different learning outcomes [333]. The overlap between IVR, HBSI and learning is depicted in Fig. 1.2.

The approach underpinning this research is that by using the evidenced link between sensorimotor-enabled learning and language learning, it is possible to understand the potential impact of HBSI in IVR. If HBSI in IVR provides learning advantages over non-HBSI interactions, then at least some form of

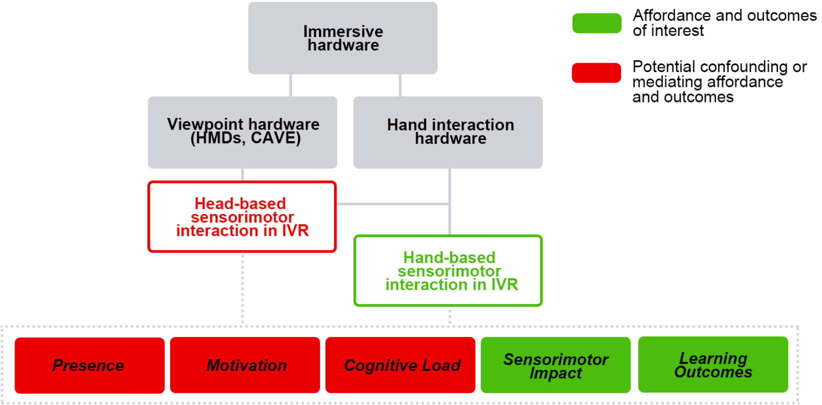


Figure 1.1: This figure shows the relationship between immersive hardware, immersive virtual reality, its predominant interaction modalities and major cognitive user outcomes. It shows (in red) the interaction (head-based interaction) and outcomes (presence, motivation, cognitive load) that this thesis is not focused on, but may prove impactful or confounding for results. It also shows (in green) the interaction (hand-based interaction) and outcomes (impact of hand-based sensorimotor interaction, learning outcomes) that this thesis is focused on. It outlines that hand-based interaction is dependent on head-based interaction for experiencing IVR, and that studying hand-based interaction is not fully separable from the head-based interaction.

embodied affordance exists in IVR, and HBSI enables them.

It is important to note that not all HBSI activity is the same, nor should there be an expectation that different HBSI activities in IVR present similar outcomes. For example, there is some evidence that different types of HBSI in the physical world have different impacts on learning, such as the distinction between gesture-encoded and action-encoded verb learning [334]. Therefore the question is not just if there is an observable cognitive distinction between HBSI and non-HBSI, but whether more nuanced distinctions within HBSI are also detectable in IVR, and if their impacts are the same as in the physical world. Specifically, insight into gesture-encoding versus action-encoding would be particularly useful for the field of IVR research, as HBSI in IVR is often referred to in the context of ‘gesturing’, despite users more often making virtual ‘actions’.

As well as exploring distinctions within HBSI, it is also possible to investigate HBSI from a system-design perspective. IVR is an authored experience, created by a system designer who makes decisions regarding the environment’s reaction to user inputs. Therefore it is possible to deconstruct interactions in a way not possible in the physical world, and attempt to understand what HBSI-related

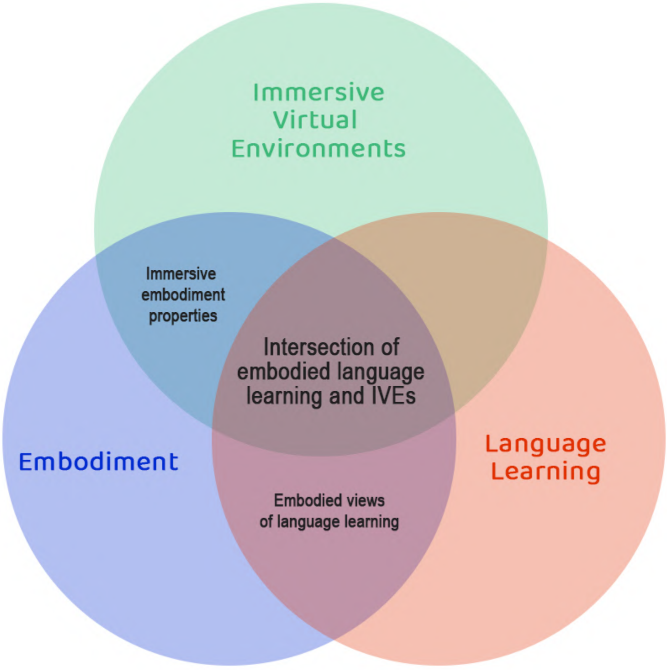


Figure 1.2: Venn diagram showing areas of relation between embodied cognition, language learning and IVR

features facilitate embodied learning affordances in IVR. In this thesis, interactional feedback is explored, in which feedback from objects being interacted with is adjusted between experiential conditions (e.g. contrasting an action in which a jug is poured and virtual water pours out, versus one in which no water is poured). This exploration helps us understand the requirements for designing the most efficient sensorimotor-based IVR learning experiences, as well as providing further insight into what aspects of HBSI have notable impacts on users (from a learning perspective), and whether replicating the physical world in IVR is the most effective system - or whether their might be ways to leverage the unique aspects of IVR for better-than-physical-world learning outcomes[12].

Finally, this research speculates on a broader question in the field of sensorimotor cognition, concerning the cognitive learning mechanisms triggered by sensorimotor-enabled learning and sensorimotor learning in IVR. Specifically, it explores how a sensorimotor-enabled interaction process causes learning gain. Is it because the sensorimotor-enabled process enables more of our bodies to interact with the learning process, allowing us to use our bodies to make meaning in an embodied way [342][183][141]? And if so, then is it because there is an innate, embodied memorisation process triggered by sensorimotor engagement [333], or is it due to the added sensory modalities that sensorimotor-based learning typically engages [229]?

Or, perhaps it is not the sensorimotor process itself that drives IVR learning, but that engaging in sensorimotor-enabled activity triggers another driver of IVR learning, such as presence or motivation. Sensorimotor interaction has been shown to be a contributor to presence [118], where presence positively effects learning [181]. It has also been shown as a contributor to motivation and engagement [184]. Of these viewpoints, the relationship between motivation and learning is the most developed. It is widely accepted that learner motivation has a positive impact on learning outcomes [66]. IVR has been recorded as being a motivating learning arena (although it is still unclear whether this can be attributed purely to a technology novelty effect [60]), and therefore it is not clear if the motivational benefits would continue to occur when IVR is as ubiquitous as other forms of computer-aided learning.

Advocates of the presence and learning relationship believe that enhanced 'presence' improves learning outcomes. The reasons for the impact of presence on learning are debated: some believe that presence alone is a phenomena that directly effects learning [217], while others believe it is a useful way of measuring how a system contributes to a variety of established variables that benefit learning, such as motivation or engagement [274][175].

Examining the learning outcomes alongside co-variables, such as the potentially confounding or mediating factors like motivation and presence, gives us wider insights into how HBSI in IVR impacts learning outcomes, and what cognitive mechanisms may be causing them.

Exploring these questions - if HBSI impacts learning, how HBSI exists in IVR, and the mechanisms behind the impact of HBSI - is important fundamental research for both understanding the cognitive impact of sensorimotor inputs into IVR, and for understanding how we can design the most effective IVR language learning and learning systems; an outcome with notable wider implications. These answers will also allow us to speculate on how humans cognitively contextualise sensorimoror interactions in IVR; and what this means in relation to wider embodied based theories of cognition.

1.2 Research questions

This thesis explores and provides evidence towards answers for the following research questions:

1. Does HBSI in IVR (i.e. manipulating virtual objects and making actions using hands) increase verb memorisation than non-sensorimotor-enabled IVR (i.e. not using hands)? (Chapter 4.1)
2. Are the verb memorisation benefits of HBSI, as evidenced in the physical world, replicated in IVR? (Chapter 4.1)
3. Are the benefits of HBSI for learning in IVR based upon enhanced embodiment directly, or is embodiment a mediating factor for another learning-enhancing property, such as presence or motivation? (Chapter 4.1)
4. How do learners cognitively experience HBSI in IVR - is it limited to gestures, or should we consider virtual actions similarly to physical actions? (Chapter 4.2)
5. Does the amount of system-feedback provided in response to HBSI in IVR have an impact on learning? (Chapter 4.3)
6. Do our findings support or refute sensorimotor-embodied theories of cognition? (Chapter 4.3)

1.3 Contributions

This work contributes towards the understanding and contextualisation of the cognitive impact of sensorimotor interaction in IVR, particularly HBSI. It also contributes to the improvement of IVR-based learning interaction design, specifically for language learning. It makes the following novel contributions to existing immersive virtual reality, embodied cognition, and computer-aided language learning literature:

1. **Leveraging HBSI in IVR for learning can lead to improved content memorisation.** This research provides evidence that learning that involves HBSI in IVR can provide learning benefits over learning that does not involve HBSI in IVR. Previously, there have been few direct comparisons between HBSI in IVR and therefore little evidence that sensorimotor interaction benefits for learning would continue to present in IVR.
2. **HBSI interaction in IVR produces similar cognitive outcomes as sensorimotor interaction in the physical world.** The learning benefits uncovered in this thesis for non-sensorimotor-interaction, hand-based sensorimotor gesture, and HBSI that involve objects manipulations (actions), match similar, previously uncovered findings in the physical world. This is evidence that HBSI in IVR and the physical world produce similar cognitive outcomes - at least when it comes to learning.
3. **The learning benefits of HBSI in IVR are not mediated by presence or motivation, and therefore likely stem from an enhanced experience of embodiment cognition.** There has been some speculation as to whether the learning benefit of sensorimotor interaction stems from the enhanced presence or motivation it might cause. However, this thesis presents no evidence that there is a mediating relationship between presence or motivation, HBSI and learning outcome. This suggests that it is the sensorimotor activity itself that is causing enhanced learning gain - a popular aspect of embodied cognition theory.
4. **The cognitive outcome of HBSI in IVR is dependent on the activity in the IVR, rather than just the sensorimotor activation occurring.** This research shows different learning outcomes between learning with HBSI that does not involved objects (gestures) and interaction that does (actions). This suggests that it is the nature of the sensorimotor interaction that is important, rather than just that the body is being activated.
5. **The cognitive (memorisation) effects of HBSI with objects is not benefited by the richness of the object's feedback.** This research found no positive memorisation effect of object interactional feedback (sound, visuals) on learning outcome. There was some evidence that the object's feedback harmed memorisation outcomes. This is evidence for the view that it is HBSI in IVR that is important, and the richness

of the feedback of the interaction is less impactful. This finding provides support for sensorimotor encoding theories of memorisation, rather than a richness-of-memory trace perspective.

6. **Evidence that leveraging sensorimotor activity increases verb learning; and the distinction of actions and gestures.** This thesis provides further evidence to the well-explored relationship between sensorimotor activity and verb learning, as well as support for the less established distinction between actions and gestures on verb learning.

1.4 Associated publications

Portions of the work detailed in this thesis have been presented in international scholarly publications, as follows:

- Chapter 3.1: Methodology, results and analysis published in **Sensorimotor learning in immersive virtual reality: a scoping literature review, IEEE International Conference on Artificial Intelligence and Virtual Reality, November 15 - 17, 2021, Taichung, Taiwan** [259].
- Chapter 3.2: Methodology, results and analysis published in **Extended Reality (XR) remote research: a survey of drawbacks and opportunities, 2021 ACM CHI Virtual Conference on Human Factors in Computing Systems, May 8 - 13, 2021, Yokohama, Japan** [254].
- Chapter 3.2: Methodology, results and analysis expanded on, and aspects of the experiment methodology of Chapter 4.2 and Chapter 4.1, in review for **The potential of remote XR experimentation: defining benefits and limitations through expert survey and case study, Remote XR User Studies Special Issue, Frontiers in Computer Science Human-Media Interaction, 2022**.
- Chapter 4.1: Methodology, learning outcome-related results and analysis published in **Evidence for embodied cognition in immersive virtual environments using a second language learning environment, 2020 IEEE Conference on Games, August 24 - 27, 2020, Kindai, Japan** [255].
- Chapter 4.1: Methodology, presence, motivation and HBSI-related results and analysis published in **Presence, embodied interaction and motivation: distinct learning phenomena in an immersive virtual environment, Proceedings of the 28th ACM International**

Conference on Multimedia, October 12 - 16, 2020, Seattle, USA [257].

- Chapter 4.2: Methodology, results and analysis published in **Actions, not gestures: contextualising embodied controller interactions in immersive virtual reality, 27th ACM Symposium on Virtual Reality Software and Technology, December 8 - 10, 2021, Osaka, Japan** [253].
- Chapter 4.3: Methodology, results and analysis (accepted, pending publishing) for **Rich virtual feedback from sensorimotor interaction may harm, not help, learning in immersive virtual reality, 28th ACM Symposium on Virtual Reality Software and Technology, November 29 - December 1, 2022, Tsukuba, Japan**.

1.5 Thesis outline

Chapter 2 A background and definition of views of embodied and sensorimotor cognition, as well as their applications for learning, language learning, and learning in IVR. Also, a background on IVR, an exploration of previous IVR applications for learning, a discussion regarding the potential embodiment afforded by IVRs, and how IVR research is conducted.

Chapter 3 The methodologies underpinning (a) the systematic process used for the scoping aspect of the background literature, (b) conducting IVR experiments remotely, and (c) the justifications for our approaches to measuring learning gain in all experiments.

Chapter 4 The methodologies and results from three experiments into the impact of HBSI in IVR on cognition.

In the first experiment (Chapter 4.1), evidence for HBSI aiding language learning; for how HBSI is its own impactful factor and not solely mediated by presence or motivation; and an anaylsis of qualitative feedback relating to potential best-practice designs for creating an IVR learning environment or experiment; are presented.

In the second experiment (Chapter 4.2), evidence for a distinction between gesture and action in IVR is presented.

In the third experiment (Chapter 4.3), a small distinction between actions with feedback and actions without feedback is presented.

Chapter 5 A discussion of the findings of all three experiments, and speculation as to their extended meaning for sensorimotor embodiment in IVR and general sensorimotor theories of cognition. Also outlines future work in the area that build upon, or cover gaps in, the evidence and approaches of this thesis.

Chapter 2

Background

This research explores if HBSI in IVR enables embodied interaction affordances; whether it is possible to measure the outcome of those affordances; and whether the outcome of those embodied interaction affordances are similar to those found in physical world. It does this through a series of comparative experiments investigating the impact of HBSI and language memorisation, a link that is well-established in the physical world. It also measures and controls for potential confounding factors related to both learning and IVR, such as presence and motivation.

Because of its broad exploration, this research requires contextualisation from a variety of areas, including an overview of popular perspectives on embodied cognition theory (2.1); embodied learning research and experimentation (2.2) second language acquisition theory to understand how embodied principles impact second language learning (2.2.2); research into the affordances of IVR, including presence and embodied interaction (2.3); an understanding of what other cognitive factors are commonly discussed in relation to IVR and IVR-based learning (2.4); an understanding of processes for IVR experimentation (2.5); and the nuances and impacts of the immersive hardware that can enable them (2.6). Each of these areas are discussed in turn in this chapter.

2.1 Embodied cognition theory

Embodied cognition is, broadly, the idea that cognitive processes are deeply rooted in the body's interactions with the world [343]. The scope of embodied cognition is still being explored, with much diversity in both its conceptions and claims [340]. It has been considered a research program [289], a thesis [340], and a grouping of distinct perspectives [343]; and multiple summative perspectives of the concept have been proposed [289][340][343].

This section provides an overview of three summative perspectives of embodied cognition, and summarises them into two meta-categorisations that allow for an exploration of embodied cognition as a design-factor within an IVR framework. Later sections explore these aspects of embodied cognition in respect to general learning outcomes, second language learning, and IVR.

2.1.1 Existing views of embodied cognition

While it is difficult to provide an overarching summary of all the potential meanings ascribed to the term embodied cognition, there have been influential summaries that have attempted to rationalise the many perspectives on embodied cognition. Three of the most well-cited summaries are outlined below:

M. Wilson's six views

Perhaps the most influential and well-cited attempt to rationalise the divergent threads of embodied cognition theory is M. Wilson's *Six Views of Embodied Cognition*[343]. M. Wilson summarises six of the prevailing "views" of embodied cognition, and addresses their credence. These views, briefly, are:

1. **Cognition is situated:** cognition takes place in the context of an environment and so must involve perception and action inside that environment.
2. **Cognition is time-pressured:** cognition should be considered not just in spatial dimensions, but temporal ones, and therefore under the pressures of real-time environmental interaction.
3. **Cognition can be off-loaded onto the environment:** cognition uses the environment to hold or manipulate information for us.
4. **The environment is part of the cognitive system:** cognition involves such a deep flow of information between the mind and the environments that the mind cannot be studied in isolation.

5. **Cognition is for action:** cognitive mechanisms (such as perception and memory) must be understood in terms of their contribution to situation-appropriate behavior.
6. **Off-line cognition is body based:** cognition is grounded in mechanisms that evolved for interaction with the environment, even when decoupled from that environment.

Of these, Wilson concludes that the fourth (the environment is part of the cognitive system) is deeply problematic, and that the sixth (off-line cognition is body based) is the most powerful.

Shapiro's three themes

Another influential summary of embodied cognition comes from Shapiro [289], who presents three themes of embodied cognition:

1. **Conceptualization:** cognitive conceptions are limited by the properties of an organism's body, so different bodies lead to difference in how the world is understood
2. **Replacement:** cognitive process are continuous, and not discrete; and an organism's body in interaction with its environment replaces the need for representational processes thought to have been at the core of cognition
3. **Constitution:** cognition relies on the body or the world as a constitutive aspect, rather than a causal aspect.

A. Wilson & Foglia's three sub-theses

There are also three sub-theses presented by A. Wilson & Foglia [344]:

1. **Body as Constraint:** cognition, both in content and nature, is significantly constrained by an agent's available bodily functions
2. **Body as Distributor:** cognition is distributed between neural and non-neural structures via the bodily functions
3. **Body as Regulator:** cognitive activity over space and time is regulated by the body, which makes cognition and action tightly coordinated.

Meta-summary: sensorimotor and environmentally-situated

Examining the three perspectives, it is possible to meta-summarise them into two broad categories or aspects of embodied cognition, as outlined in Table

2.1. These meta-categories are based on Beer's attempt to interpret embodied principles into a functional system, which presents an *embodied* nervous system, that depends on the properties of having a body for sensing and actions; and a body that is *situated*, in that it exists as part of an environment [19].

1. **Sensorimotor Embodied:** M. Wilson (#2, #5, #6), Shapiro (#1, #2, #3) and A. Wilson & Foglia (#1, #2 ,#3) relate cognition with interactive processes such as actions; time- and function-constraints; and limited or deeply related to the affordances of the body.
2. **Environmentally Situated:** M. Wilson (#1, #3, #4) and Shapiro (#3) and A. Wilson & Foglia (#2) relate to the importance of cognition leveraging or existing within an environment.

Meta-category	M. Wilson	Shapiro	A. Wilson/Foglia
Sensorimotor	Cognition is for action Cognition is time-pressured Off-line cognition is body based	Replacement Conceptualization Constitution	Body as Regulator Body as Constraint Body as Distributor
Situated	Cognition is situated Cognition off-loaded onto environment Environment part of cognitive system	Constitution	Body as Distributor

Table 2.1: Meta-categories for the different embodied cognition summaries

These meta-categories, although broad, provide guidance for exploring notable aspects of embodied cognition in ways applicable to IVR. While they may not perfectly cover all the meanings ascribed to the term embodied cognition, it allows us to explore some of the most important ones: understanding the impact of sensorimotor action-sequences and exploring bodily affordances, and investigating environment contributions. We will discuss these meta-categories in turn.

2.1.2 Sensorimotor embodied cognition

In M. Wilson's, Shapiro's and A. Wilson & Foglia's summaries presented above, sensorimotor embodied cognition is discussed from five perspectives:

1. Action is the primary motivation for various cognitive activities
2. Perception and action are fundamentally inseparable in lived cognition
3. Sensorimotor systems constitute part of a singular system of cognition that involves the environment, body and nervous system

4. Sensorimotor actions and perceptions are underpinned by the affordances of our bodies
5. Mental structures originally evolved for perception or action, now provide higher level thought-processes

The motivation for the first perspective is the belief that key cognitive functions developed in order to allow creatures to take part in adaptive activity[88]. This approach is founded on the idea that vision evolved for the purpose of allowing guided actions such as reaching and grasping [104], and memory "evolved in service of perception and action in a three-dimensional environment"[95]. Glenberg even goes as far as to argue that the traditional approach to memory as 'for memorising' should be replaced by "the encoding of patterns of possible physical interaction with a three-dimensional world".

These beliefs are backed by links found in brain activity investigations. For example, visual input of an action being carried out, can, in the viewer, activate areas of the brain responsible for carrying out those actions[109].

There is also evidence that the action-affordance of an object impacts cognitive responses. In one study, response times were fastest when the response hand was congruent with the hand that should be used to grap the object (for example, if a tea-pot had a handle on the left-hand side, responses were faster when the left hand was used)[323]. This theory is reinforced by the observed body-object interaction effect (BOI), an effect which shows that whether or not a human body could physically interact with an object that a word refers to effects cognition concerning the word. For example, recent research has shown that words for high-BOI objects (ones that are easier to interact with, e.g. fork) are recognised faster and recall is less erroneous than responses to words for low-BOI objects (e.g., mountain)[158]. Further studies about actions, gestures and learning outcome are discussed later in this chapter (see 2.2).

M. Wilson also presents an argument for a more "indirect, flexible, and sophisticated" relationship with the action-affordances of objects, based not just on how an object is currently presented, but on how "information about the nature of the external world is stored for future use without strong commitments on what that future use might be." [343] Knowledge of objects are stored with a rich information about the potential action affordances: a piano can be used to make music, block doorways, or be smashed for firewood.

The second perspective differs slightly from the first in that perception is not considered subservient to action, but sits with action as fundamentally inseparable in lived cognition [329]. From this view, the contents of perception are determined by the actions an organism takes, and the actions an organism takes are guided by its perceptions of the world. This is a deeper relationship between sensing and motor action than in the first perspective. However, both the first and second perspectives can be summarised as the belief that interactivity, and potential for interactivity, impacts cognition. The third perspective extends these actions into the environment. It suggests that actions serve as a composite part of a holistic system in which “the environment, body, and nervous system are each dynamical systems and are in continuous interaction”[19]. Therefore cognition uses the environment, the body and the nervous system at once, and actions have a cognitive impact as both an interaction channel between the body and the environment, as well as of themselves.

Evidence for this perspective is often presented from research into gestures. Research into spatial reasoning in five-year-olds found that they described distinct spatial rotation strategies when gesturing as opposed to when using speech. This suggests that gestures presented one method of thinking about the problem, while speech expressed a another[76], despite both being environmental interactions.

The fourth perspective is the idea that our sensorimotor actions and perceptions are underpinned by the affordances of our bodies, or in Wilson & Folio’s words, “an agent’s body functions to significantly constrain the nature and content of the representations processed by that agent’s cognitive system”. They argue that this inherently makes some forms of cognition easier or harder (or even impossible) because of a creature’s bodily characteristics.

Shapiro refers to (something very similar to) this as Conceptualization, and presents the following arguments for it:

- Concepts are embodied.
- Thus, concepts are constituted in part by activity in the perceptual, emotion, and motor areas of the brain.
- Differences in embodiment cause different kinds of activity in the perceptual, emotion, and motor areas of the brain.
- Hence, differently embodied organisms will possess different concepts.
- Therefore, differently embodied organisms will think differently

Shapiro presents evidence for this through Kaschak and Glenberg[98], who found that subjects have greater difficulty understanding sentences that describe actions to which human bodies are not suited. This poses interesting questions for embodied cognition in virtual spaces, where the interactional options and aspects embodiment are distinct from the real-world; and especially for homuncular flexibility [346], in which participants control additional virtual limbs that do not exist in the physical environment.

The fifth perspective is that mental structures, which originally evolved for perception or action, have been “co-opted to run off-line” and provide higher level thought-processes[343]. The processes have become abstracted from the physical inputs and outputs of their original function, so much so that they can appear as entirely ‘cognitive’ functions. This view has many proponents [95][307]). Wilson presents research backing for this viewpoint, citing the following explorations in memory research:

- Working memory has “separate storage components for verbal and for visuospatial information”[10], and thus working memory off-loads information onto sensorimotor systems in the brain.
- Episodic memory, which are a class of memories defined by records of spatiotemporally localized events, as experienced by the remember.
- Implicit memory (or skill learning) automisation, where practice allows new skills to become automatized, reducing cognitive load and circumventing the representational bottleneck[80].

Further evidence for off-line, body-based cognition comes from cognitive linguistics, which proposes that linguistic syntax is tied to semantics, such as image schemas or gestures representing embodied knowledge of the physical world [174][315].

These five perspectives describe different aspects of embodied cognition as it relates to sensorimotor embodiment. While they may differ in how they explain the impact of embodiment on cognition, they broadly agree on one key aspect: the sensorimotor apparatus of the body can be leveraged to impact cognition.

Aside from what constitutes embodied cognition, there is another question that permeates embodied cognition discourse: is cognition only influenced by action sequences and body and environmental states (what Shapiro calls a “causal” relationship[289])? Or is there a comprehensive brain-body-environment cognitive systems that radically departs from traditional “mentalist” views of cognition science (what Shapiro calls a “constitutive” aspect)[198]? This thesis does not explore this question.

2.1.3 Environmentally-situated embodied cognition

The three summaries above present two views of the role of the environment and the body’s ‘situated’ nature:

- An environment or representations of an affect impact cognition
- Cognition ‘extends’ into the environment

For (1), the view is that a type of cognition, situated cognition, takes place in the context of task-relevant inputs and outputs, which occur in an environment or situation [343]. This view is not generally considered controversial (e.g., [53]; [58]).

It is believed we can reduce the cognitive workload by making use of the environment in strategic ways: putting information into the world to be accessed as-needed rather than taking the energy to memorised it, and altering the environment to reduce additional cognitive work [164]. For example, Kirsh and Maglio [164] reported a study involving Tetris, in which players who actually rotated and moved objects had better outcomes than those who imaged the rotations or movements.

Wilson argues that this kind of situatedness is useful for spatial tasks in particular, but limits the ideas range of applicability as a cognitive strategy. Cognitive tasks, such as planning, remembering, and day-dreaming, do not fall into this situated category[343]. However, Wilson also presents an argument for situating these tasks using environmental activities: drawing Venn diagrams, doing math with pencil and paper enable physical manipulation that saves cognitive work that would otherwise be done in our brains. Another approach is imagining situated settings to aid memory storage, such as the memory place remembering technique [133]. This adds spatial dimensions to abstract things that are trying to be memorised.

Regardless of these, literature stresses using the world as "its own best model"[35] when designing for embodied cognition. Rather than attempt to mentally store and manipulate all the relevant details about a situation, we can store and manipulate those details in the situation itself.

The second perspective suggests that we do not just use environments to aid cognition, but that cognition extends into our environments, and that enables types of cognition we would not otherwise be able to do [59]. An useful example comes from Clark:

> When we are busy writing and thinking at the same time [... it] is not always that fully formed thoughts get committed to paper. Rather, the paper provides a medium in which, this time via some kind of coupled neural-scribbling-reading unfolding, we are enabled to explore ways of thinking that might otherwise be unavailable to us.

Clark argues that it is not just that the environment is being leverage to aid cognition (as in the previous view), but that the pen and paper allow the brain to cognise in otherwise impossible ways. Evidence for this is presented in the methods that human beings take to organize their surrounding environments to ‘ease’ cognitive burdens, such as placing keys by doors to remember their location or cataloging files alphabetically to minimize searching demands.

These two perspectives have implications for the use of environmentally-situatedness on learning: the first perspective suggests that just existing in an environment is enough to enhance cognition. To leverage the second, however, you also need to ensure an environment provides adequate affordances for extended cognition.

2.1.4 Summary

It is clear that embodied cognition is a developing area with a variety of constitute aspects, each with implications for the study of cognition. The perspectives of M. Wilson, Shapiro and A. Wilson & Folia’s may have some distinctions, but they universally suggest that there is a strong link between sensorimotor activities, the environment, and cognition.

The meta-categorisations created by delineating M. Wilson, Shapiro and A. Wilson & Folia’s embodied cognition views into the taxonomies of sensorimotor embodied cognition and environmentally-situated cognition, suggests two perspectives from which to explore embodied affordances of IVR. While both are intriguing investigations, this thesis will only explore sensorimotor embodied cognition. The next section discusses how sensorimotor embodied approaches have been evidenced to impact learning outcomes in the physical world.

2.2 Sensorimotor embodied cognition and learning outcomes

Sensorimotor-enabled embodied cognition and its effect on learning processes have been described in literature for almost 100 years [72]. Contemporary research has many examples of experiments where a group learning in a highly embodied scenario out-performs a lower embodied control across a variety of learning subjects, including language [97][96][94][149][2][193][192], STEM subjects [228][62][208][11][28][1][143] and learning related-skills [48]. Despite this ubiquity, there is also still some evidence that sensorimotor-embodied learning approaches can be detrimental to learning outcomes [84][248].

In this section, evidence is presented for and against leveraging sensorimotor embodied cognition for enhancing learning outcomes for language learning. First, experimental research into the impact of sensorimotor activity on learning, and the theories underpinning this research, is discussed. Then, approaches and methods commonly used in language learning are discussed.

2.2.1 Experiments and theories

Language learning has been one of the predominant subjects of sensorimotor-enabled experimental research. These studies are based on the perspective that language in the brain is represented by a sensorimotor network formed from experiences collected on the concept being memorised [250]. Further evidence for this is presented in investigations that show that concrete words are easier to remember than abstract ones [36]; as they are richer in sensorimotor stimulus than abstract words.

Experimental evidence

There have been numerous investigations into links between sensorimotor activity and both language and language learning. In part, this has been in opposition to the previously predominant cognitivist perspective of language [192]. A meta-analysis [140] of the field supported the hypothesis that the motor system is activated during language comprehension, with brain imaging studies often demonstrating a link between reading words and activity in parts of the brain associated with engaging in that action. For example, reading action words (like kick or throw) activate areas of the brain responsible for carrying out those actions [116], while reading odor words (like the names of spices) prompt olfactory-related brain activity [102].

Investigations setup to explore a link between sensorimotor activity and second language learning have generally found positive results. For reading, children “acting-out” scenes being read with toys showed recalled 33% more information compared to children who had objects present but were not allowed to manipulate them [97]. An abstracted version of this study, using computer images of objects found a similar-sized effect [96][94]. For writing, it has been demonstrated that physically writing (instead of typing, a less motor-involved action) improved the processes of letter recognition, naming, and letter composition (as well as reading comprehension)[149].

Actions and gesture

Sensorimotor activity can take many forms, and activity that involves using the body but not interacting with another object (gesturing) also appears to play an important role in memorisation. Of the multiple gesture types, iconic gestures (gestures similar to the action, meaning or idea they represent, e.g. gesturing a throw to request a ball be thrown) have been considered “fundamental to all languages ... [bridging] the gap between linguistic form and human experience” [321]. Areas of the brain responsible for iconic gestures and physical actions have been shown to activate when associated words are used or heard [193], [210]. For language acquisition, iconic gestures are considered universally important for both first and second language acquisition [230], [355], and have been considered an additional “mode of thinking” [213] for second language learners.

Controlled experiments have shown that word learning that was encoded with gestures was recalled better than learning that was encoded without gestures, or where gestures were presented only during testing [2]. Similar gesture versus non-gesture encoding experiments have found similar results [193][192], even when the presentation of gestures was mediated via a television screen. Further research has shown that the encoding gestures need to be contextually congruent with the words being encoded [155][193], and that learning for words suffered when encoded with incongruent actions (such as kicking to learn the word ‘punch’).

Computer-aided language learning has also seen benefits from using action and gesture. Edge [74] found users enacting a sequence of movements to complete a foreign-language movement instruction performed better than a control; Macedonia [196] had participants imitate a pedagogical agent’s gestures and visually learn words accompanied by gestures; and Repetto [264] found that when recognizing novel words, participants made less errors for words encoded with gestures compared to words encoded with pictures.

Theories: embodied or multi-modal

There is evidence that combining embodied actions, like gesture, with spoken production causes enhanced language memorisation. Gesture and spoken production work together to enhance communication, forming an "an integrated system in language comprehension" [156] with demonstrable benefits in word understanding when gesture and speech are congruent. Growth Point Theory [214] hypothesizes that speech and gesture interact and influence one another throughout the planning and speaking of utterances, with gestures helping speakers to "internalise the abstract via the concrete". However, proponents of the multi-modal theory of cognition [229] might argue that, as learning benefits arise when more modes of interaction and feedback are involved, that the benefits of sensorimotor activity for learning stem from the added modalities of the sensoritmotor activation.

Experimentally, Kelly demonstrated positive Japanese memorisation outcomes by having learners combine gesture with simultaneous, relevant spoken production [155]. Later, Bergmann and Macedonia achieved the same but with sentence learning, rather than singular words [21]. Both of these studies showed that when a learner used gesture with spoken production they achieved better learner outcomes than spoken production alone. Interestingly, these contradict the original findings of the Total Physical Response language teaching approach, which demonstrated that students' success when attempting to learn both listening and speaking together was significantly decreased [7]. It remains to be discovered how these two modes relate to language learning in IVR.

2.2.2 Sensorimotor language teaching approaches

In order to investigate how sensorimotor-enabled IVR impacts language learning; or what language learning can tell us about embodied cognition in IVR; it is both important and useful to look at the rich vein of applied linguistic theory regarding second language learning and sensorimotor and environmental embodiment. This chapter presents popular second language teaching theories, approaches and techniques related to this embodied language acquisition theory within a sensorimotor embodied cognition framework.

In a second language tuition, there are three teaching methods that leverage this sensorimotor relationship: Total Physical Response, Task-based Language Teaching (TBTL) and Communicative Language Teaching (CLT).

Total Physical Response

Of Wilson's six claims of embodied cognition [342], the claim that off-line cognition is body-based - "when decoupled from the environment, the activity of the mind is grounded in the mechanisms that evolved for interaction with the environment" - is particularly relevant to the relationship between gestures and language acquisition. The use of off-line embodiment was operationalised by applied linguists for second language acquisition some decades before embodied cognition theorists began to coalesce around the theory, in the form of the Total Physical Response [7] teaching approach. Asher found that learners of Japanese performed significantly better at recognising spoken words if they performed an action related to the word while learning.

There have been attempts to explain the benefits afforded by embodied actions and gestures for language acquisition outside of embodied cognition. Asher noted that the learning benefits of his approach could be explained through increased learner motivation, while later studies found any light to moderate physical activity during encoding - such as performing actions - is beneficial to vocabulary acquisition and retention [279]. However, there is strong evidence that the positive relationship between iconic gestures and acquisition is not entirely mediated by physical activity or higher motivated. Experiments have shown that iconic gestures relevant to the words being encoded (e.g. jumping while learning the word for 'jump'), rather than unrelated gestures (e.g. jumping to learn the word 'kick'), have significant retention benefits [350], [193]. If the learning benefits were solely caused by the enhanced motivation provided by learning with physical activity, or merely the effect of the physical activity itself, it would be difficult to explain why the use of related gestures was superior to unrelated ones.

Further evidence for the unique encoding potential of iconic gestures for language learning is found in Macedonia's work [193], which showed that word acquisition related to iconic gestures activated different parts of the brain than word learning with unrelated gestures. The former activates areas associated with the pre-motor cortices that control bodily movement, while the latter activates areas associated with cognitive control.

Whatever the reason for the benefits of using embodied actions or iconic gesture as a tool for language memorisation, experimental results in embodied controls and computer-aided language learning have proved positive: Vasquez [332] used iconic gestures to help with listening skills related to verbs that correspond to the gesture enacted by the learner; Edge [74] had users enact a sequence of movements to complete a foreign-language movement instruction; Macedonia [196] had participants imitate a pedagogical agent's gestures and visually learn words accompanied by gestures; and Repetto [264] found that when recognizing novel words, participants made less errors for words encoded with gestures compared to words encoded with pictures.

Task-based Learning

Task-based language teaching (TBTL) is method that refers to “the use of tasks as the core unit of instruction in language teaching” [265]. In this context, tasks have been defined as “an activity in which a person engages in order to attain an object, and which necessitates the use of language” [327]. An example is a painting task in which the learner has to use second-language colour names to request different paint pigments to complete the painting activity. In this sense, TBLT promotes language goals in terms of language use, rather than linguistic content [327], and prescribes both activity and agency to learners [223]. These are clearly aspects that fit closely with the embodied sensorimotor concepts discussed earlier.

Communicative Language Teaching

Communicative Language Teaching (CLT) is the actioned teaching approach of communicative competence (CC). CC has evolved since its original definition [134][42][45], but broadly refers to either a curriculum of knowledge that extends beyond just language [134], and as an approach to language acquisition “guided and evaluated by the learner’s ability to communicate” [277]. CLT differs from traditional language teaching in that it is concerned with not just knowledge of the language itself, but also the ability to, and appropriacy of, using it in a social and cultural context [134], including in interpersonal discourse [44]. The components of CLT are interrelated, and thus it is considered optimal to teach language amongst social and cultural context, through discursive interaction [44].

A CLT curriculum is equally defined by what it is not: it excludes (or deprioritises) explicit grammar instructor [320], the conscious study of rules of language systems, which include “phonology, syntax, morphology, lexis, semantics, pragmatics, discourse” [218]. This does not mean CC is unconcerned with the importance of grammatical accuracy, but that accuracy is encouraged through communicative practice rather than a learner’s structural awareness of the language (as evidenced in [276]). Therefore, CLT is a strong match for sensorimotor embodiment (and also environmental embodiment), as both are based on language as an input/output process and interactional tool [44].

2.2.3 Summary

The breadth of research into the impact of sensorimotor approaches to learning have mostly demonstrated positive learning results. However, an explanation for how and why this occurs is yet to be agreed on, with many varying perspectives, some of which (multi-modal theory) do not particularly involve sensorimotor embodiment or embodied cognition. It is also still unclear how types of sensorimotor activity, such as gestures as opposed to actions, may contribute to learning - especially in IVR, in which the entire environment is a simulation.

2.3 IVR affordances: presence and embodied interaction

Immersive virtual reality (IVR), as defined in this thesis, is a computer-mediated experience in which the user is perceptually surrounded by a virtual environment [189]. The virtual environment is a 3D world in which users can interact or navigate to some extent. Research suggests that this perceptual surrounding stimulates cognitive and psychological responses in users, such as feelings of presence (place illusion and plausibility [295]). This has been referred to as the first affordance of IVR [141].

Many types of IVR allow HBSI through immersive hardware. Typically, this takes the form of motion controllers or hand tracking, these approaches allow users to manipulate and interact with the IVR using the physical body's hand positions and actions. It has been suggested that this embodied interactivity should be considered the second affordance of IVR [141].

In this section, we provide an overview of both of IVR's hypothesised key affordances - presence and embodied interaction.

2.3.1 Presence

The term 'presence' is widely used in IVR literature. However, there are a large number of divergent and overlapping definitions of the concept [188], as outlined in Lombard and Jones' summary:

- "We define presence as the feeling of being located in a perceptible external world around the self." [338]
- "Presence is the experience of being engaged by the representations of a virtual world" [137]
- "Presence is defined formally as the perceptual illusion of nonmediation" [187]
- "Presence is tantamount to successfully supported action in the environment" [352]
- "The sense of presence considered here is... a numinous [i.e., supernatural, sacred, holy] sense of otherness" [51]

These separate definitions agree on three common features of presence: (1) presence requires involvement into the virtual environment; (2) presence is defined as a subjective experience; and (3) presence is a multidimensional construct [92]. None of the definitions limit presence to being an experience restricted to interactions with IVR. As such, two IVR-specific sub-sections of presence have been established: place illusion and plausibility [297].

Place illusion

One of the most referred to definitions of presence is the "sense of being there" in a virtual environment, despite being physically situated in the physical environment [345]. Even back in 1995, with less advanced immersive hardware, it was suggested that IVR had crossed a "psychological threshold, a point at which our perceptual systems are so immersed in the simulation that the user already begins to feel some of the sense of being there, the early flushes of a powerful presence" [24]. This specific definition is also known as spatial presence [22] or place illusion [295].

Much research on 'presence' is predominantly concerned with studying place illusion [22]. However, as a subjective experience, there is no way to directly measure it [295]. Assessments are typically based on questionnaires regarding physiological and behavioural responses [295]. Based on these, studies have demonstrated that IVR generates stronger place illusion than non-immersive virtual reality, such as the type experienced via desktop, mouse and keyboard combinations [204].

Research shows that place illusion is linked to other cognitive and psychological factors. For example, increased place illusion has been linked with more [70][110] and less learning [204]; although it is sometimes unclear if researchers are measuring presence as place illusion or as one of the wider definitions. Place illusion does not need HBSI to occur; a user can look around an environment without any feedback from the system (except for responding to gaze direction) and feel place illusion [295].

Plausibility

Plausibility is the "illusion that what is apparently happening is really happening (even though you know for sure that it is not)" [295]. Slater's example is if a "virtual human approaches and smiles at you, and you find yourself smiling back, even though too late you may say to yourself – why did I smile back, there is no one there?". Plausibility has been demonstrated experimentally in IVR through proprioceptive drift in a virtualised version of the rubber-hand experiment [296].

It is believed that the plausibility of IVR is closely tied to the experience's internal consistency and the extent it meets users' expectations [293]. The user expectations are formed from a combination of their knowledge of the real world (external plausibility), as well as what they know about the virtual environment (internal plausibility) [123]. For example, it is externally plausible that a gun fires bullets, but it could be internally plausible that a gun fires flowers. If, however, the gun predominantly fires bullets and on one occasion fires a flower, the internal consistency might be broken (especially if the user's expectations were for a serious police drama, and not something more akin to Alice's Adventures in Wonderland). If a gun were to fire a flower in a police drama setting, then the user's sense of plausibility could be broken due to the implausibility of that occurrence in that setting [295], and may not re-form [93].

It is believed that place illusion and plausibility can lead to realistic behaviour in IVR [295]. Early experiments show that differences in levels of plausibility (between 'high' and 'low' plausibility) do not impact the experience of place illusion, suggesting they are not dependent phenomena[123].

2.3.2 Embodied interaction

Over the last three decades, embodied interaction has been referred to as a second major affordance of IVR [309][336] with particular relevance to research on learning [141][203].

Steur refers to IVR interactivity as "the extent to which users can participate in modifying the form and content of a virtual reality environment in real time" [309]. While this definition accommodates interactions from a HMD, either via head locomotion or eye gaze, it is more often considered as referring to hand-based interaction. This is because that hand-based activity and gesture are considered to play a special role in human interaction; "kinesthetically activat[ing] larger portions of the sensorimotor system and motoric pre-planning pathways than [eye and vocal interaction] and gesture may lead to stronger memory traces" [100]. As Johnson-Glenberg explains, "the ability to control movement via gaze is one form of agency, but the ability to control and manipulate objects in the 3D environment is perhaps a different and deeper form of agency with many more degrees of freedom" [141].

The potential impact of embodied interaction in IVR on users is not fully clear, as many of the existing comparison studies have not isolated immersion and interactivity [203]. However, one model of the relationship between presence, embodied interaction and user outcomes, CAMIL, has the affordance causing a positive effect on nine variables: feelings of agency; physical presence; intrinsic motivation; self-efficacy; extraneous cognitive load interaction; extraneous cognitive load environment; situational interest; embodied learning; and learning [242].

In testing the CAMIL model, Makransky and Petersen found that physical presence and agency arise from IVR embodied interactions, which then influence learning via the affective and cognitive factors of situational interest (becoming more interested in the learning topic) and embodied learning (the unique learning properties offered by learning in an embodied way). However, their work found that the use of embodied learning decreased knowledge acquisition, a result which Makransky and Peteresen attributed to a lack of congruency between bodily actions and learning content in their study [242].

In summary, embodied interaction in IVR has been considered an important aspect of the IVR experience, but its impacts are only now starting to be explored and understood.

2.4 IVR, learning and sensorimotor-enabled learning

The previous sections have discussed the embodiment-related affordances of IVR, embodied cognition, sensorimotor approaches to learning and language learning. In this section, we summarise literature that has previously brought some or all these topics together, and explore the differing theoretical explanations for the impact of sensorimotor activity on learning inside IVR.

Previous systematic reviews and meta-analyses into IVR and learning provide some useful definitions of different learning-related scopes of IVR studies. These were scopes of IVR hardware (from CAVEs, to three degree-of-freedom HMDs, to six-degree-of-freedom HMDs); scopes of learning (including vocational training, cognitive and education, skill-improvement, empathy enhancement); and scopes of sensorimotor engagement.

The broadest IVR learning meta-analysis, in terms of scope of IVR, learning and embodied interaction, is Howard's meta-analysis of virtual reality hardware and software for personal development [128]. It found no evidence that input hardware, which includes but is not limited to, sensorimotor inputs, had a significant effect on cognitive development. Howard presents four potential explanations for this: (1) input hardware may have little impact on mechanisms that may subsequently influence learning outcomes; (2) current input hardware influences important mechanisms, but input hardware may not influence these mechanisms enough; (3) input hardware may influence mechanisms that have little effect on IVR intervention outcomes compared with other, more important IVR aspects; and (4) specialized input hardware may not influence the nature of tasks enough to incur a substantive effect on outcome.

If we subscribe to the belief that embodied cognition could play a role in learning development whatever the medium, then these questions can be re-summarised as: does existing IVR hardware and design enable embodied cognition in a form that provides comparable-or-better learning outcomes to physical-word embodied approaches?

A potential limitation in applying Howard's findings to the study of embodied sensorimotor interaction in IVR is that his scope of IVR and learning were both broad. The IVR studies he examined took a variety of less embodied and sensorimotor-enabled forms, including CAVEs or three degree-of-freedom HMDs; and his definition of cognitive learning included both academic subjects, vocational training and other forms of personal development.

Therefore it might be important, when specifically meta-analysing sensorimotor activity in IVR, to avoid such broad explorations. Helpfully, there are some other reviews and meta-analyses approached IVR learning with more tightly defined scopes. These reviews also provide useful observations around learning subject matter, which will be discussed below.

2.4.1 Topics of IVR learning

A common distinction in existing IVR learning reviews is between academic education and vocational training [49][251][73][138]. Checa and Bustillo [49] noted that this was often reflected in how studies were being evaluated: the majority of studies evaluating training focused on task performance, whereas the education-dominated ones evaluated knowledge acquisition. This approach suggests that difference scopes of learning should be, and often are, treated differently. There were also distinctions in the types of academic education, with one reviews exploring studies of IVR learning in high-education contexts [251], with another focused on K-12 and higher educational subject-learning contexts [73]. Jensen & Konradsen's review found six studies, out of 21 they discussed, that specifically investigate academic education rather than vocational training [138].

2.4.2 Sensorimotor IVR learning

The impact of sensorimotor engagement in IVR educational learning has not been thoroughly investigated. Jensen & Konradsen identified five studies that leveraged sensorimotor input [138], but these were all concerned with skills acquisition - such as juggling - rather than cognitive learning. The review concluded that the prevailing question was not if HMDs should be used, but rather how and for what should HMDs be used. They also theorised that IVR learning could work successfully with educational approaches and theories such as Constructivism and active learning.

Natale et al. do not specifically mention sensorimotor engagement, but do discuss an idea similar to that of embodiment [73]. They conclude their study by mentioning that the possibility for users to feel present in IVR, to use their bodies in a natural way, and to live sensory experience similar to those in the real world could be promising as advanced learning instructional strategies, and call for further research into how these aspects could affect the learning affordances compared to less immersive ones.

Radianti et al. presented a thorough summary of the main ideas behind the existing IVR learning paradigms, including Behaviorism, Cognitivism, Constructivism, Experientialism, Connectivism, but do not specifically refer to a sensorimotor approach [251].

There is, however, a lack of investigation into the learning possibilities and benefits afforded by leveraging sensorimotor techniques in IVR learning, especially for academic learning. While Howard touches on the area with 'input hardware', his search is more interested in the hardware itself rather than its application in the learning context. Similarly, for Radianti et al. it is not mentioned in their list of identified learning paradigms. Only Natale et al. refers to the possibilities of sensorimotor-based approaches, although this is in the limitations section of their own research.

Immersively, a Mixed Reality physics learning study saw embodied students improved their performance by 76% on the second trial compared with 51% for those who used the simulation without bodily cues [1]; and in an IVR experiment, learning was significantly better for users who had to move their body than not [143]. Additionally, students in the "high embodiment" condition retained their learning better when retested a week later.

However, this lack of investigation, coupled with the growing number of investigations into sensorimotor IVR academic learning, suggests a dedicated review is needed.

2.4.3 Confounds of sensorimotor IVR learning

Contrasting views have attributed the learning benefits noticed when leveraging embodiment in IVR to four main factors. These are the relationship between between embodied IVR, embodiment and learning [183]; embodied IVR, motivation and learning [201]; between embodied IVR, presence and learning [216]; and between IVR, cognitive load, and learning [204].

While many IVR learning investigations monitor motivation, presence or embodiment, few have explored these factors in the same experience, or investigated how they relate to each other to promote learning (or if they interact at all). There is increasing demand for this kind of fundamental understanding of what factors influence learning in IVR [128], especially if we are to understand the specific role of embodiment in IVR.

IVR: embodied cognition

The embodied cognition perspective is that neither presence nor motivation are the key to the benefits offered by sensorimotor interactions. Instead, if cognitive processes are rooted in the body's interactions with the world[343], so by replicating more naturalistic interaction through sensorimotor controls, we can enhance learning by synthesising a more natural learning process [196].

Sadly, many comparative investigations in this area are of limited value for this discussion as there are confounding differences between the embodied and control environments, or confounding factors are not monitored. For example, interactional richness is often added to sensorimotor controls but restricted from their non-embodied interfaces; such as in Johnson and Glenberg's learning comparison of the dynamic embodied creation of 3D models with a simple text entry system [141]. Similarly, other experiences monitor learning differences but not motivation or presence changes [332][349].

IVR: motivation

Literature concerning motivation is extensive, with the phenomena having been studied from multiple perspectives, resulting in many theoretical frameworks. Broadly, motivation is considered the energisation and direction of behavior [245]. Strong links between motivation and learning have been found, with the phenomena considered the "key to persistence and to learning that lasts"[52], with many reviews showing evidence for a strong correlation between motivation and learning success [66] [146].

These links are also well-evidenced in instructional games [326] [236], with games primarily seen as a means to enhance intrinsic motivation [121]. A learner who is intrinsically motivated undertakes an activity "for its own sake, for the enjoyment it provides, the learning it permits, or the feelings of accomplishment it evokes" [178].

IVR has been noted to present motivational benefits compared with less immersive environments [201][55][238]. Motivation in IVR has predominantly been studied from a learning perspective, with IVR seen as a means to enhance intrinsic motivation [201] to an extent even greater than non-immersive games[121].

It is still unclear whether the motivational uplift is due to the novelty effect [60] of using IVR, and therefore if the motivational benefits will continue to occur when IVR is as ubiquitous as other forms of computer-aided learning. However, early research has given evidence that this is not the case [130].

IVR: presence

Advocates of the IVR, presence and learning relationship believe that enhanced 'presence' (the feeling of "being there" in a virtual environment [300]) improves learning outcomes. The reasons for the impact of presence on learning are debated: some believe that presence alone is a phenomena that directly effects learning [217][241], while others believe it is a useful way of measuring how a system contributes to a variety of established variables that benefit learning, such as motivation or engagement [274] [175].

Research into whether presence independently affects learning, and the mechanisms responsible for its impact, has thus far failed to prove conclusive. Although it appears high levels of presence among learners are related to better learning outcomes [150], there are also studies that show the opposite: increased presence correlated with worse learning outcomes [204]. Presence has also been difficult to define and measure consistently, with presence levels in IVEs varying wildly [187].

One possible explanation for this is that the enhanced emotional involvement of feeling ‘present’ in a situation [41] encourages better learning. This would explain why IVR concerned with emotive subjects, such education around climate change, show both increased presence and learning [207]. The link between presence and strong emotional responses in users, such as empathy and anxiety, is well-established [283].

However, perhaps the most prolific explanation is that the learning benefits offered by increased presence are a result of a positive relationship between presence and motivation [201] [204] [274] [175]. In essence, more presence means greater motivation, which means better learning.

A more prosaic perspective, not investigated here, is that presence has no causal relationship with learning, and that there only appears to be one due to the affordances of the immersive hardware that enable both presence and learning [128]. For example, it is far more difficult for a learner to get distracted from learning when using a head-mounted display, as the screen is strapped to their face [128] which, coincidentally, also serves to increase their sense of presence.

IVR: cognitive load

Cognitive Load Theory suggests that extraneous cognitive load interferes with the perception of experiences; and high element interactivity can cause this[313]. A similar view is found in the Cognitive Theory of Multimedia Learning[211], where it is suggested that more modes of media interaction can cause cognitive issues. As IVR is considered a highly interactive medium that offers rich multimodal interaction, researchers have begun investigating links between IVR, a user’s cognitive load, and potential effects this might have. Studies have found that immersion causes more presence but creates higher cognitive load, ultimately hurting learning outcomes [204]. The question of whether extraneous processing is something that IVR experiences will be forever burdened with, or it is merely a product of unfamiliarity with the system or poor environment design, remains to be answered.

Where found, lower learning rates in immersive environments have sometimes been attributed to issues with cognitive load [280] [328] [204], with claims that virtual immersion creates a large cognitive load that detracts from a learner’s ability to memorise information. As embodied controls are considered to increase immersion, then according to the above, they should also increase cognitive load. However, Steed et al. found that the use of embodied controls in IVR (when paired with a tracked self-avatar presentation in the virtual environment) actually reduced cognitive load [306]. This suggests that immersion stemming from sensorimotor controls could be different to other types of immersion and have a different cognitive impact. If Steed is correct, we would see a reduced impact on cognitive load from this study.

Summary

In summary, it is clear that to provide more insight into the causes of learning benefits in IVR, further experimental research is needed that allows for a control of embodiment factor (such as allowing HBSI and not allowing it), and which monitors presence, motivation, cognitive load and learning gain. In this way, it may be possible to begin to understand if embodied cognition via sensorimotor embodiment contributes to the success of learning in IVR, or whether it is mediated by another factor.

2.5 Conducting IVR experiments and learning experiments

Using IVR for experimental research presents unique opportunities and challenges for researchers compared with non-immersive virtual environment or non-virtual environment studies. The nature of the hardware means that, in lab settings, it could be more susceptible to viral and bacterial transmission between participants (e.g. Covid-19). However, as consumer immersive hardware market penetration increases, there could also be an opportunity for remote immersive studies, in which the entire experimental environment is contained within the HMD experience. These benefits and challenges of IVR, the potential solution of remove experimentation, are discussed below.

2.5.1 IVR experiments

Benefits and challenges of IVR experiments

There are many suggested benefits to using IVR as a research tool: it allows researchers to *control the mundane-realism trade-off* [5] and thus increase the extent to which an experiment is similar to situations encountered in everyday life without sacrificing experimental control [27]; create *powerful sensory illusions within a controlled environment (particularly in VR)*, such as illusions of self-motion and influence the proprioceptive sense [301]; *improve replication* [27] by making it easier to recreate entire experimental environments; and allow *representative samples*[27] to experience otherwise inaccessible environments, when paired with useful distribution and recruitment networks.

Many of the challenges faced by IVR are similar to those found in non-immersive virtual worlds [233]. These include the challenge of ensuring the *experimental design* is relevant for each technology and subject area; ensuring a consistent feeling of *self embodiment* to ensure engaged performance; avoid *uncanny valley*, in which characters which look nearly-but-not-quite human are judged as uncanny and are aversive for participants; *simulation sickness* and nausea during VR experiences; *cognitive load* which may harm participation results through over-stimulation [204]; *novelty effects* of new technology interfering with results [78]; and *ethics*, especially where experiences in IVEs could lead to change in participants' behaviour and attitude in their real life [14] and create false memories [285].

Covid-19 considerations: remote IVR experiments

This research was disrupted by the COVID-19 pandemic, beginning 2020. The HMDs used for accessing IVR were considered a potential virus transmission vector. As such, there was a need to examine remote experimentation techniques to adhere to social distancing requirements, as well as general health and safety and ethical concerns. There has been little research into remote IVR experimentation. In terms of research outcome, Mottelson & Hornbæk [222] directly compared in-lab and remote, out-of-lab IVR experiment results. They found that that while the differences in performance between the in-lab and remote study were substantial, there were no significant differences between effects of experimental conditions. Similarly, Huber and Gajos explored uncompensated and unsupervised IVE samples and were able to replicate key results from the original studies, but with smaller effect sizes paper [131]. Finally, Steed et al. showed that collecting data in the wild is feasible for IVR systems [304].

Ma et al. [191] is perhaps the first published research on recruiting remote participants for IVR research. The study, published in 2018, used the Amazon Mechanical Turk (AMT) crowd-sourcing platform, and received 439 submissions over a 13-day period, of which 242 were eligible. The participant demographics did not differ significantly from previously reported demographics of AMT populations in terms of age, gender, and household income. The notable difference was that the IVE research had a higher percentage of U.S.-based workers compared to others. The study also provides insight into how remote IVE studies take place: 98% of participants took part at home, in living rooms (24%), bedrooms (18%), and home offices (18%). Participants were typically alone (84%) or in the presence of one (14%) or two other people (2%). Participants reported having "enough space to walk around" (81%) or "run around (10%)". Only 6% reported that their physical space would limit their movement.

While Ma et al.'s work is promising in terms of reaching a representative sample and the environment in which participants take part in experiments, it suggests a difficulty in recruiting participants with high-end VR systems, which allow six-degrees of freedom (the ability to track user movement in real space) and leverage embodied controllers (e.g. Oculus Rift, HTC Vice). Only 18 (7%) of eligible responses had a high-end VR system. A similar paucity of high-end VR equipment was found by Mottelson & Hornbæk [222], in which 1.4% of crowdworkers had access to these devices (compared to 4.5% for low-end devices, and 83.4% for Android smartphones). This problem is compounded if we consider Steed et al.'s finding that only 15% of participants provide completed sets of data [304].

An alternative approach to recruiting participants is to create experiments inside existing communities of IVE users, such as inside the widely-used VR Chat software [271]. This allows the research to enter into an already large community of active users, rather than trying to establish one themselves. There are significant limitations of building experiments in platforms not designed for experimentation, such as the ability to communicate with outside services for data storage, and absence bespoke hardware interfaces.

The above led us to conduct the experiments outlined in further work via remote experimentation procedures.

2.6 Immersive hardware and IVR affordances

The technological hardware that allows users to experience IVR is typically referred to as immersive hardware. There are many divergent incarnations of immersive hardware, and there is some evidence that variations in the design of immersive hardware impacts users' experiences of IVR, and the outcomes they have from it.

For example, hardware attributes such as tracking level, stereoscopy, and field-of-view (FOV) have been found to have an impact on place illusion [69]. The different hardware choices can lead to issues regarding research reliability and validity [331], especially for meta-analyses, in what has been referred to as a research "wild west" [25].

In an attempt to clarify some distinctions in immersive hardware, and to justify the choice of immersive hardware used in this thesis, four hardware distinctions are discussed here: between CAVE and HMD hardware, between 3DoF and 6DoF interaction, between types of commercially available types of 6DoF HMDs; and, importantly, between HBSI-enabled and non-HBSI systems.

2.6.1 Experiential differences between immersive hardware

CAVE vs HMD

There are two preeminent hardware approaches for enabling IVR. CAVE systems, which surround users with multimedia hardware such as screens, projections and speakers [68]; and head-mounted displays (HMDs), which attach screens and speakers to a headset worn by users, and also detect the orientation of the user's head to simulate a viewpoint in a virtual world [15].

The two approaches have been compared numerous times over the last two decades. Some comparative studies have found significant benefits to using HMDs over CAVE systems, suggesting HMDs lead to faster task completion times [65] or induce more significant therapeutic movements, emotional responses and increased immersion [77]. Others have found that CAVEs induce a higher level of presence and elicited more anxiety in exposure therapy [145], induced "a greater user experience ... with significant difference in presence, engagement, flow, skill, judgement and experience consequence" [318], and generally have better performance outcomes [163]. And others have found "similar levels of engagement, engrossment, immersion, experience rating" between the two [185], no significant differences concerning player performance and immersion [30], similar "user responses to thematic relations such as engagement, embodiment,and preference" [244], and no significant differences in effectiveness in an exposure therapy session [215][18]. These results are summarised in

Table 2.2.

Author	Year	Property	Better Hardware
Cordeil	2016	faster task completion time	HMD
Elor	2020	therapeutic and emotional responses, immersion	HMD
Juan	2009	presence and effectiveness in exposure therapy	CAVE
Tcha-Tokey	2017	presence, engagement, flow, skill, judgement	CAVE
Kim	2012	general better performance	CAVE
Liu	2019	engagement, engrossment, immersion, experience	Both
Bowman	2001	player performance and immersion	Both
Pilpot	2017	engagement, embodiment,and preference	Both
Meyerbroker	2011	effectiveness in exposure therapy	Both
Baus	2011	effectiveness in exposure therapy	Both

Table 2.2: CAVE and HMD comparative study results

Much of this comparative research may be somewhat limited in relevance for modern immersive hardware, however, given advances over the last two decades. For example, FOV differences were once quite dramatic, with CAVEs offering FOV up-to levels found in the human; while HMDs used from 2001 and earlier ranged from 30° to 80° (according to a summary of the field in [166]). Contemporary HMDs have FOVs above 100° [138], making them much closer to those found in the human eye.

Perhaps the major distinction in CAVE and HMD hardware is in HBSI with the systems. CAVE systems have struggled with allowing users to interact with the virtual environments around them, with interaction devices used in CAVES being labeled as “generally awkward and spatially inaccurate” [117], particularly as the hand-based action occurring on the user, but the feedback from the system occurring at some distance away on the surrounding screen. HBSI interaction in HMDs, when available, present a virtualised representation of the hands, which can also interact directly with objects in the virtual world.

Whatever the differences between CAVEs and HMDs, the commercial availability and affordability of consumer-grade HMDs means that this is by far the more popular vector for experiencing IVR, with estimates suggesting that 43.5 million 6DoF HMDs will be sold by 2025 [317]. Therefore, using HMDs instead of CAVEs seems like a more practical research approach for understanding IVR.

HMD 3DoF vs 6DoF

HMD systems can be further distinguished by the opportunities for interaction that they provide. One notable distinction is how user head movements are integrated into IVR. There are two prevailing approaches, categorised as three degrees of freedom (3DoF) and six degrees of freedom (6DoF). While both allow the user's head movement as a former of sensorimotor interaction with the environment, the former only tracks the orientation of the user's head, while the latter also tracks where the user's head moves in the physical world. The experiential distinction is that, with 3DoF, the user can look in any direction but is fixed into a single virtual location. With 6DoF, the user may augment their virtual viewpoint by moving towards or away from objects, or leaning to undertake new perspectives.

There are limited direct comparisons between 3DoF and 6DoF systems, possibly because the commercial virtual reality market has been quick to adopt full 6DoF. However, there is evidence that 6DoF offers higher immersion, engagement and a higher sense of presence than 3DoF [246][290][46]. Additionally, qualitatively, participants have been recorded as preferring experiences in which they have ability to move their viewpoint in a virtual reality space, as opposed to being fixed in place [290].

The generalisability of the above studies are somewhat limited by their experimental design, however. All mentioned studies compared different models of headsets, features of which could have proved influential beyond the 3DoF/6DoF variable. And one study allowed HBSI for the 6DoF setup and not the 3DoF [246], which could have influenced immersion.

Distinctions between commercial HMDs

There are a limited number of different HMDs available, but there is still evidence of distinctions in the user experience and outcomes between different devices.

Significant differences have been found between two of the most popular consumer IVR HMDs (the HTC Vive and Oculus Rift), with the HTC Vive leading to better user effectiveness and a more usable system in a locomotion-by-teleportation task [206] and in a pick-and-place task [312]. The HTC Vive has also been found to offer a larger working range (7m) than the Oculus Rift (4.25 m), although marginally worse tracking accuracy [29].

Modifications to individual types of headset can also influence outcomes. Researchers were able to influence a user's ability to estimate distance [38] of the inside the Oculus Rift DK1 HMD by reducing the field-of-view and adding weight to the headset, which resulted in an increased chance of users underestimating distances.

These results suggest that the design of HMDs impacts various user cognitive or psychological responses. However, comparative research on the impact of design distinctions in consumer HMDs is lacking (and given the frequent release cycle of HMD updates, potentially fruitless). Nether-the-less, it is important to understand that HMD choice, available degrees of freedom and distinctions between HMD and CAVE can all impact experiment results, despite all being grouped into the IVR canon.

HBSI-enabling hardware

IVR is increasingly, although not exclusively, paired with highly embodied controls that allow HBSI [142]. These tools are considered important in creating "natural user interfaces", and are often referred to by a myriad of names: embodied controllers; motion controllers; synced hand controls; hand controls; gesture controllers; kinesthetic controllers; touch controllers; or wands. The use of these tools allows for user inputs that result in HBSI that are often labeled as either inputs; gestures; actions; interactions; embodied actions; multi-modal interactions; or kinesthetic actions.

There are multiple approaches to enabling HBSI in IVR, which create further distinctions in how an IVR is experienced and its potential impact on its users. Typical approaches detect bodies, hands, hand controllers, and use hardware such as lasers, visible spectrum cameras, depth-based multi-camera system, IR-based cameras, gloves or wrist-bands.

Broadly, HMDs, rather than CAVEs, have been considered more powerful in establishing embodiment, as they enable the ability to see virtualised body parts as they interact with the virtual environment [141]. Research has suggested that users find these kind of multi-modal interaction methods to be "natural and immersive" [136]. However, there is limited research into the cognitive impact of different embodied controllers hardware approaches. There is also evidence that highly embodied experiences (using HBSI to do contextually-relevant actions) create more presence than the low embodied ones (no embodied controllers) [148].

For HBSI hardware, users may either manipulate controllers or have their body movements recorded and virtualised by cameras or lasers. Direct comparisons between controller-based and hand-based IVR interactions are limited, but one study found that physical controllers outperformed the hand-and-gesture controls in subject-perceived accuracy, efficiency, and satisfaction, although with no significant difference in task completion time[20]. Similarly, another found that hand controls produced more errors than a type of embodied controller, as well as taking longer and, where significance was found, was less favourable [83].

2.6.2 Summary

In summary, IVR hardware is varied and these variations could have profound implications for research outcomes. The distinctions in different types of HMDs, hardware enabling HBSI, between 3DoF and 6DoF, between HMD and CAVE systems, make it clear that investigations into IVR are a broad church that could contain many influential variables. What these variables are, what they affect and to what degree is currently unknown; and based on the regular updates to modern HMD devices, may never be studied in a way that has meaning.

However, it is clear that the sensorimotor affordances can be deeply different between IVR systems, which could affect IVR user outcomes. This section serves as a caution regarding the generalisability of IVR studies and their applicability across different types hardware, and offers guidance towards picking the best-fit solution for a study of HBSI in IVR: 6DoF HMDs.

Therefore for this research, 6DoF HMDs are used. To enable HBSI, the environments force the use of controllers rather than camera-based hand tracking, even on devices where this is an option (i.e. Oculus/Meta Quest 2). For the first experiment, the Oculus Rift-S is used; while for the remote experiments, both Oculus Touch and HTC Wand controllers are supported (although almost all participants used the Oculus Touch system).

2.7 Literature summary

This literature review has presented existing perspectives and experimental research around IVR and the embodied interaction affordances it offers. In the process, it has highlighted the theoretical underpinnings of such an affordance existing by examining theories behind embodied cognition, as well as presenting evidence of previous research in the area of sensorimotor learning in IVR.

The review has provided justifications for using learning and language learning as a subject for exploring the sensorimotor interaction experience of IVR, by presenting an overview of the embodied theories of language acquisition and various experimental evidence.

It also discussed potential confounding factors to be examined or accounted for in the design of experiments to explore the sensorimotor affordances of IVR, such as presence, motivation and cognitive load, and hardware and experimental factors, such as different types of immersive hardware.

Chapter 3

Methodologies and approaches

This chapter presents and justifies three methodologies that underpin core aspects of this thesis. These include the systematic process used for the exploration of existing sensorimotor-enabled IVR experimental research; an expert survey into the approaches, advantages and drawbacks of conducting IVR experiments remotely; and current practices for examining language learning in experimental settings.

It has is divided as follows:

- Chapter 3.1: The justification, systematic process and findings of a scoping review concerning experimental research into sensorimotor-enabled learning in IVR, which aided in the creation of the background chapter and provided guidance in later experiment design.
- Chapter 3.2: The justification, research approach and findings (advantages, limitations and areas of note) regarding conducting experimental IVR research remotely, rather than in-lab.
- Chapter 3.3: An outline of current practice to inform how we should examine and test language learning based on existing experimental research.

3.1 Scoping sensorimotor-enabled learning in IVR: method and results

3.1.1 Introduction

In the previous chapter, numerous reviews into IVR and learning were presented [128] [138][251][73][49]. However, there was an absence of attempts to specifically synthesize the divergent approaches and interests of sensorimotor-enabled IVR learning interventions in order to present an understanding of the state-of-the-art of sensorimotor-enabled IVR experimentation.

A review exploring sensorimotor-enabled IVR learning is important, both to avoid potential over-generalisations that may be present in less specific IVR and learning examinations [128] and to understand any research approches and specific user outcomes afforded by sensorimotor-enabled IVR experiences. These include (1) which learning topics are being explored; (2) what type of sensorimotor approaches are used, and how, and what type of hardware is enabling them; (3) what theoretical justifications or motivations are behind these decisions; (4) what experimental process are used and how is learning gain being measured; (5) what learning results have been uncovered; and (6) what other experiential measures are also being monitored.

This systematic search identified 14 documents reporting on experimental sensorimotor-enabled IVR learning studies, which universally presented positive learning results. However, it also identifies questions around differing definitions of learning gain, potential confounds (such as situation and context); and an absence of longitudinal or holistic learning studies, of sensorimotor-orientated embodiment questions, of discussion of how different IVR hardware might impact results, and of connections between sensorimotor research. This section provides a useful context for the current state-of-the-art of sensorimotor-enabled IVR learning research.

3.1.2 Scoping review methodology

The scoping review was conducted using an approach informed by Xiao and Watson [348] and consists of three stages: (1) formulating the research problem and developing the review protocol; (2) searching the literature, screening for inclusion, extracting data, analyzing and synthesizing data; and (3) reporting the results.

As this was a scoping review, and so attempted to discuss the breadth of the field, there was no requirement to formally assess paper quality [348]), however all included papers had passed a formal peer-review.

Formulating the research problem and developing the review protocol

The initial approach was to scope the field of experimental research concerning learning using sensorimotor techniques in IVR, reviewing the efficacy, experimental methodologies, theoretical approaches and hardware-use related to sensorimotor-led learning experiments inside IVRs.

The goal was to provide a complete overview of experimental research into sensorimotor-led learning inside IVR in an attempt to identify a conceptual boundaries of the field, the size of the pool of research, types of available evidence, and any research gaps.

However, during pre-mapping the scope was reduced from all learning to only academic (or non-vocational, non-skill-based) learning.

Pre-mapping

In order to understand if the area of investigation was feasible, a pre-mapping approach [32] was used to identify potential subtopics within the proposed research area. To explore this broad area, Google Scholar was used with the following search term:

'virtual reality' OR 'virtual environment' OR 'virtual simulation' OR 'vr' AND 'immersive' AND 'learning' AND 'sensorimotor'

As expected, Google Scholar provided many results (n = 5600). An inductive analysis was performed on the first 250 abstracts, attempting to uncover any immediate themes that might be relevant to the review process. As a result, an important taxonomy of subtopics was defined: type of learning. The types of 'learning' returned were categorises as cognitive (or semantic, knowledge-based learning); vocational or skill-training; physical or mental recovery; and affective exercises (e.g. encouraging empathy or reducing anxiety). These categories are similar to those identified in Jensen & Konradsen's review [138] of the IVR learning space, and map to the definitions and distinctions between academic and training found in other reviews.

To reduce the size of the investigation, and to align with our research interests, the scope was limited to experiments concerned with cognitive learning.

Research question development

The research questions for this review were informed by criteria from previous systematic literature reviews related to IVR and learning. In Feng et al.'s [82] IVR serious games systematic literature review, they determined two primary types of research question: pedagogical impact, concerned with learning outcomes and measures; and the behavioral impact, concerned with behavioural outcomes and measures. Similarly, Natale's [73] review tracks the following outcomes and measures: learning measures, learning result, effect size (learning outcome-based factors); and motivation measures, motivation result, effect size (behavioral impact measures).

Feng et al.'s review included further question specific to their topic: "what are the essential elements for developing IVR serious games?", with five subquestions relevant their review: what teaching methods, navigation solutions, senses simulated, narrative methods, and NPC contributions are used in experiments.

To add specificity for sensorimotor-enabled interaction into this review, special attention was paid to the following sensorimotor-related factors: the type of embodied controller used; the embodiment conditions the study was comparing between; whether it was one-off or across multiple sessions; whether it was a standalone study or part of a pedagogical process; what topics were chosen; approaches and justifications that researchers used; how data was record recorded; and what IVR-related experiential measures were used.

These were summarised into the following core research questions:

- What areas of study (topics) are being explored for sensorimotor-enabled cognitive learning in IVRs?
- What theoretical learning or teaching approaches are being used to justify these approaches, and are they dependent on the topic?
- How are these studies being conducted?
- How is (or what type of) sensorimotor-enabled activity being used?
- What hardware is being used to enable this activity?
- What objective learning results (efficacy) have these studies demonstrated?
- What subjective learning experiences have these studies demonstrated?

Developing the review protocol

Selection criteria

In order to include experimental research concerning cognitive learning using sensorimotor-enabled interaction in IVRs, the following criteria for paper inclusion were outlined:

- Use immersive virtual reality (not just virtual reality)

- Explore cognitive, semantic or knowledge-based learning (exclude studies on vocational training, skill-development, physical and mental recovery, and affective changes like anxiety or empathy)
- Designed to explore relationship between sensorimotor activity and outcomes
- Use a HMD-enabled immersive environment (not CAVE)
- Use controllers that enable some form of bodily input with the environment and learning situation beyond the user interface menus, including physical and camera-based methods
- Be experimental or quasi-experimental
- Report either objective (learning change) or subjective (experiential) measures
- Full text is accessible and available
- Full text is in English

Search strategy

The strategy sought to discover a comprehensive selection of peer-reviewed research documenting experimental or quasi-experimental studies related to the topic. Five research databases were identified that would be a good fit for this research: two interdisciplinary (SCOPUS, Web of Science), and others in computer science and engineering (IEEE Xplore), education (ERIC) and psychology (PsycINFO).

Prior reviews of learning in IVR have limited the time-period of their analysis in order to control for the introduction of "a new generation of HMDs [offering] a better quality user experience" [138]. As high field-of-view HMD IVR systems with embodied controls were not widely available prior to the release of modern commercial headsets, such as the HTC Vive and Oculus Rift; results were limited to those published in 2016 or after. CAVE systems were also not included due to user interaction devices being labeled as "generally awkward and spatially inaccurate" [117]; and to align the research with the prevailing trend in IVR research towards the use of consumer-grade HMDs.

Each database was searched in March 2021 with keywords based on this Boolean search string:

('virtual reality' OR 'virtual environment' OR 'virtual simulation' OR 'vr' OR

'head-mounted display' OR 'immersive environment') AND ('learning' OR 'training' OR 'education') AND ('sensorimotor' OR 'kinematic' OR 'embodied')

Papers that passed the inclusion criteria were then snowballed exhaustively in both directions, exploring papers from their reference list and using Google Scholar to identify papers that had referenced them.

Data collection

The database listings and search results were archived for record keeping, reproducibility, and crosschecking [165].

Each paper's abstract was examined to determine if it matched the above selection criteria, and if it was unclear, the conclusion section also read (as per guidelines from Brereton [32]). If it was still unclear, studies were included for further study. The reason for exclusion was recorded for each paper.

3.1.3 Searching the literature, screening for inclusion, extracting data, analyzing and synthesizing data

Searching the literature and screening for inclusion

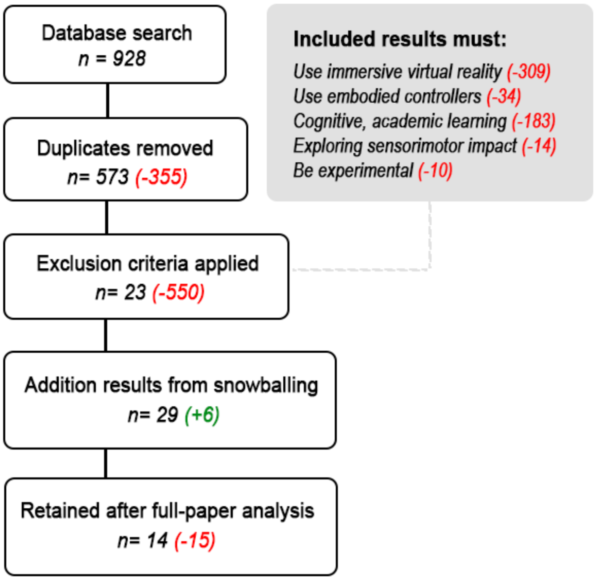


Figure 3.1: Flowchart depicting the article selection process

The search uncovered 928 results (ERIC, 18; PsycINFO, 30; IEEE Xplore, 118; SCOPUS, 470; Web of Science, 292). Of these, 573 remained after removing duplicates. The following removals occurred at each stage of our criteria: [A] Must use immersive virtual reality (not just virtual reality), must use a HMD-enabled immersive environment (not CAVE) (309 removals). [B] Use controls that prioritise sensorimotor input as part of the learning process (and so behold controller or keyboard button-pressing) (34 removals). [C] Explore cognitive, semantic or knowledge-based learning (not studies on skill-development, physical and mental recovery, and affective changes like anxiety or empathy) (183 removals). [D] Designed to explore relationship between sensorimotor activity and outcomes (14 removals). [E] Be experimental or quasi-experimental and report either objective (learning change) or experiential measures (10 removals).

This left 23 papers. References were snowballed backwards and forwards and identified a further six papers. This left 29 papers for a full reading for inclusion. At this stage, we removed a further 15 papers due to the following three categories: the paper was reporting on the same experiment as another included paper; there were not recorded experimental results; or the target was not cognitive learning. This left 14 eligible papers [107][287][270][23][316][47][257][237][50][139] [115][91][200][332]. This process is illustrated in Fig. 3.1, and the list of included papers can be found in the appendix.

Extracting, analyzing and synthesizing data

An inductive approach to categorisation was used, which resulted in the research questions listed above. Responses were thematically analysed using an inductive approach based upon Braun and Clarke's six phases of analysis[31].

3.1.4 Results

Topics

The predominant learning topic investigated was second language education (n = 7); specifically word memorisation (6). Of the papers, five included noun memorisation and two included verb memorisation.

There was a diverse, if limited, range of other topics present in the papers: biology (2), mathematics (1), physics (1), computer science (1) and geography (1).

There was also a single investigation into a sensorimotor-led cognitive skill development that could provide benefits across multiple learning subjects: perspective taking, or "the ability to mentally represent a [visual] viewpoint different from one's own".

Theoretical approaches

All but one of the papers referenced a theoretical approach behind their motivation to investigate sensorimotor-enabled IVR learning. The most evident theory was that of embodied cognition, which was also referred to as embodied learning and learning from embodied interaction (7). Four (of the six) investigations into language learning referenced language embodiment theories.

Three papers referred to approaches that are related to embodiment, although not sensorimotor embodiment specifically. These were (1) spatial cognition, the importance of using the environment around you for meaning making, storage and processing; (2) Constructivist, specifically Interactionist, in which interactions (but not necessarily sensorimotor ones) play an important role in learning, and (3) situated learning, which includes aspects of both spatial cognition and Interactionist perspectives.

One paper justified its approach from a motivation-perspective, suggesting that the motivation benefits of body-based interaction in IVR should cause enhanced learning outcomes.

Research motivations

There was a notable distinction between the research motivations for language learning investigations and those of other topics. Four of the language learning investigations were motivated by a desire to provide empirical evidence for an existing theory of embodied benefits to language learning; and whether these extend into IVRs.

The major justification for non-language studies was to improve on perceived shortcomings of physical world sensorimotor teaching approaches, including practicality and cost. For example, more detailed examples of these include whether a sensorimotor IVR experience could adequately replace the experience of re-constructing canine skeletons (compared with a real box of bones); whether seeing inside a hand can improve understanding of hand biology over cadaver dissection (and also remove the requirement of having access to a severed hand); or re-balance gender-related interest in Computer Science learning through a more active and interesting approach.

Three papers cited a research motivation to see if a sensorimotor approach could maintain learning whilst also improving motivation and interest in the subject.

Experiment methodologies

The studies presented five methods for measuring learning gain. The most popular approach was a pre- and post-exposure test of participants' knowledge, with the performance change between the two measures showing the learning gain (6). Other approaches examined the participants' response times to questions, with quicker responses presented as evidence of better learning outcomes. One study monitored participant movement via HMD data to understand if the user was performing an action that represented the target learning (bowing at the start of interactions as a cultural Japanese learning process). Another monitored performance in a game to determine progress of the learning material - better game performance being indicative of a better desired learning outcome. Finally, one study simply asked participants to report if they felt they had achieved better learning via the IVR system.

Eleven studies compared the sensorimotor IVR condition with a control, although the types of control varied. The most prominent comparison was with a non-sensorimotor IVR alternative (3). There were also comparisons with incongruent sensorimotor activation (2); non-IVR sensorimotor interactions (2), such as a physical skeleton arrangement task versus a virtual one; non-IVR and non-sensorimotor digital setting (2); low-interaction (flashcards); non-interaction (watch-only) and AR embodiment. One study compared non-tangible IVR sensorimotor activation with a tangible version, using a bespoke device to provide physical feedback to the virtual interaction.

Types of sensorimotor

The studies used a variety of different approaches for enabling participants to interact in a sensorimotor-embodied way. These could loosely be broken down into two categories: (1) sensorimotor interaction with objects in an environment, and (2) sensorimotor-as-input method. For the former, participants would manipulate 3D objects in a contextual setting with varying levels of feedback. For the latter, participants moved their bodies in novel ways and the system would recognise and react. A key distinction between the two being that the first involved sensorimotor interactions with virtual representations of objects, whereas the later did not.

Among studies engaging sensorimotor interaction with objects in the environment, the dominant form was allowing participants to pickup virtual objects, move them around and conduct actions with them. For example, picking up a jug and pouring it so water comes out in order to learn the verb 'to pour' or the noun 'jug'.

Two studies allowed participants to pick up and move objects, but did not allow them to be used in an interactional way with the environment. For example, a jug could be picked and moved, but if moved as if to be poured, the system did not provide an interactive outcome (e.g. water falling from the jug).

One further study allowed users to pick and move objects, but rather than provide an interaction congruent with the physical world, instead presented an outcome only possible in IVR. In this case, participants could pick up bones and move them in the air, where they stayed suspended as the user constructed a floating, gravity-less canine skeleton.

The types of activation in the sensorimotor-as-input method studies were more varied. In one, participants would move their own hand to see a digitised version (and an enlarged digitised version) move in the same way but with added visualisations, to demonstrate how the typically hidden hand tendons worked. Another, using a custom suspension-rig, allowed participants to experience a situation similar to micro-gravity in order for them to experience how the related forces work. One studied used dance-based input in order to organise a virtual computer programming experience, while another had users slash words with virtual lightsabers in order to make active selections of multiple options.

Among these types of sensorimotor implementation, there was an even split between types of head-mounted display (HTC and Oculus), with bodily inputs coming from HTC 'wands' (4), hand-tracking (2), bespoke hardware (2), camera-based full-body tracking (1), head movement (1) and Oculus 'Touch' controllers.

Learning efficacy

The conclusions of all of the examined papers supported the use of IVR sensorimotor approaches for learning. However, the criteria used to justify that support varied between studies. The criteria was either: (1) learning occurred, as opposed to no learning or negative learning; (2) learning occurred at a similar rate to the control; or (3) learning out-performed the control.

The majority of studies defined 'learning occurring' as knowledge gain based on pre- and post- test result changes. However, one study examined the speed of participant responses, with faster response speeds demonstrating improved learning. Another study measured learning by the self-reported 'learning experience', i.e. did participants feel they had learned during the experience?

There was also variance between studies around how many tests showed learning gain. Of the papers, six noted significant learning benefits in the IVR sensorimotor condition across all of their reported learning gain measures. However, two noted learning gain benefits only on some of their reported indicators of learning gain. Interestingly, these two studies were both on word memorisation, and there were contrary outcomes between them. One showed only a significant benefit of sensorimotor encoding in an immediate test, but not in a retention test taken one-week later; while the other only showed a significant in

Type	**Measure details**
Usability	2x SUS, 1x MEEGA+, 3x Custom/not given
Sickness & Comfort	1x SSQ, 1x SSQ adaptation, 3x Custom/not given
Presence	2x IPQ, 1x Nowak &Biocca, 1x Custom/not given
Motivation	1x MEEGA+, 1x Custom/notx given
Enjoyment	2x Custom/not given
Embodiment	1x Gonzalez-Franco & Peck
Task Load	1x NASA TLX
Flow	1x Custom/not given
Intrinsic Interest	1x Custom/not given
Concentration	1x Custom/not given
Satisfaction	1x Custom/not given
Enthusiasm	1x Custom/not given
Interest in subject	1x Custom/not given
Perceived Usefulness	1x Custom/not given
Perception of Time	1x Custom/not given
Fun	1x Custom/not given
Preference	1x Custom/not given
None	3x no measures presented

System Usability Survey (SUS)[179], Model for the Evaluation of Educational Games (MEEGA+)[243], Simulator Sickness Questionnaire (SSQ)([157], Igroup Presence Questionnaire (IPQ)[281], Task Load Index (NASA TLX)[114], Nowak & Biocca [227], Gonzalez-Franco [103]

Table 3.1: Showing the non-learning measurements used in the studies

a retention test taken one week later, and not in the immediate test.

Two studies noted similar results between the IVR sensorimotor condition and an alternative.

No papers reported learning gains worse than the control in all reported conditions.

Experiential measures

The studies featured a wide range of experiential measures (see Table 3.1). The most commonly used measures were sickness/comfort, usability and presence. Broadly, the measures could be categorised into the following categories: properties strongly linked to IVR (presence, embodiment, sickness); measures of the system usability (usability); general engagement properties (enjoyment, excitement, motivation, interest/subject interest, preference); and experience in the session (flow, NASA task load index, concentration).

Interestingly, there was only one study that specifically measured IVR embodiment [270], using the Gonzalez-Franco and Peck embodiment questionnaire [103]. There were no questions specifically asking about participant experiences of the sensorimotor activity.

3.1.5 Discussion

The following observations were made about different aspects of the surveyed papers, and thus the field of experimental sensorimotor-enabled IVR for learning (as captured by this search).

Conceptual consensus and confounds

Unsurprisingly, the IVR sensorimotor investigations found were predominantly grounded within embodied theoretical approaches. Perhaps more interesting are the studies that grounded their explorations in other conceptual approaches, as these could help highlight potential mediating factors for investigations looking to support or repudiate sensorimotor and embodied cognition approaches in IVR (and also the inverse).

One of these approaches was motivation [115], which as a factor in improving learning outcomes is well-established [66][146]. There is also evidence that sensorimotor learning activities increase motivation [177][311][122][182][288]. However, the idea that sensorimotor-enabled interaction in IVRs prompts motivation and therefore is the key learning driver in these interventions is unproven. Indeed, results presented in this thesis (see 4.1) show no evidence of motivation serving as a mediating factor for embodied learning - at least in a one-off session.

Another given approach was situated learning [50][139]. Despite being contextualised by their authors inside a situated cognition framework, these studies deeply deployed sensorimotor interactivity; one in which physically bowing was a key interactive input [50], while the other involved pointing-and-grabbing objects [139]. In fact, some degree of situatedness (such as realisitic and congruent learning contexts for interactions), was part of all of the surveyed IVR experiences.

The risk of situatedness being a confound for a pure sensorimotor vs. non-sensorimotor comparison was present in the majority of studies, as only some of the papers compared (congruent) sensorimotor interaction against either non-sensorimotor or non-congruent sensorimotor interaction in the same IVR [107][47][257][91].

If researchers are to truly understand the impact of sensorimotor interaction in IVR, then controlling or engaging with IVR-induced variables, especially ones that have a theoretical backing as contributing to learning outcomes (such as situatedness) is important.

Language dominance

There was a clear language dominance in the studies surveyed. The prevalence of language acquisition in these studies is likely due to the long history [6] and encouraging results [319] of using sensorimotor-enabled teaching approaches in applied linguistics outside of IVR. The ability for sensorimotor engagement to enhance word memorisation has been widely explored (and evidenced) in this field, and so exploring whether these benefits transfer into embodied virtual interactions is a logical next-step.

That said, even these studies are not close to giving us a clear understanding of the impact of sensorimotor interaction on language learning as a whole. Word memorisation is an important but subsidiary part of what is considered language teaching or second language acquisition, and researchers need to be able to move beyond these initial examinations if they are to be able to understand if and how sensorimotor engagement is beneficial for the myriad of language skills (although one example of this is included here, with the attempt to teach appropriate bowing etiquette [50]).

What is 'success'?

While all papers reported successful learning outcomes from their sensorimotor approach, the definition of what success was varied. Learning improvement over a control; learning improvement in a specific condition over a control; the same performance as a control; testing response speed; 'experience' or motivation improvements were all presented as evidence for the success of a sensorimotor approach.

For studies that use multiple measures of success, such as retention tests after a set period of time, or knowledge gain and answer speed, it is unclear how contradictory results between different tests should be used to promote or refute a learning 'success'.

While the definition of success is certainly learning intervention-specific, perhaps it would be useful for the field to define types of success or potential metrics for success, in order to allow thorough future meta-analyses of the field. This would most likely be determined by specifically defining the desired goal of the intervention, whether it be to see whether a physical world sensorimotor approach transitioned in IVR without harm, in which case the desired outcome would be parity between conditions, or if testing a variation of a sensorimotor interaction, in which a superior outcome in one condition would would be the goal.

Learning 'session' variation

The prevalence of word memorisation studies unveiled another variable that might benefit from codification and standardisation in future - the scale of the learning to be achieved in an intervention. The discrepancies between amount of learning material are evident in the language-based studies, which asked participants to learn either six[107], 12[139], 14[91] or 20 [257][115][332] words. Potentially, six words might be too few to detect a difference in learning outcomes (although in this case, response time, not accuracy, was the performance indicator), while 20 may introduce too much confusion for optimal learning performance. This difference also adds another confounding factor for cross-analysis.

Longitudinal lacking

None of the studies surveyed were longitudinal, nor involved a longer curriculum of learning. This is a known problem with experimental learning studies [98], and one that might be exasperated by the additional development requirements that sensorimotor IVR studies require. With all the papers reporting some form of learning gain from engaging these systems, however, a longitudinal study seems like a logical and pressing next-step to understand how useful IVR sensorimotor learning is for practical teaching applications.

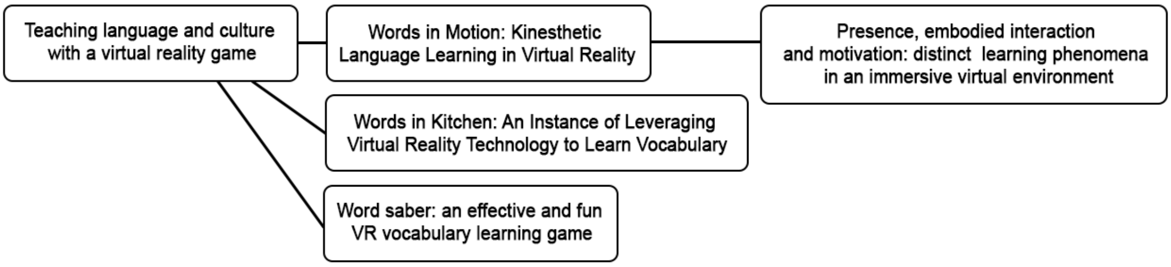


Figure 3.2: Diagram depicting the lack of referencing between papers found in review. Three papers referenced "Teaching language and culture with a virtual reality game"; one further paper referenced one of those. No other direct relationships between papers were found

Sickness and presence above others

There were a wide variety of experiential measures used across the studies. The frequency of measures of sickness and presence show that these two factors are starting to become well-established as important aspects of IVR studies. A notable absence, however, were any measurements concerning the sensorimotor activity, and only a single use of a measure of embodiment. Perhaps these studies would have benefited from surveying how embodied or sensorimotor-activated the participants felt in their sessions, and in what ways. Questionnaire-based embodied surveys for IVR are now emerging and being validated [103], although it remains to be seen if they are suitable for measuring sensorimotor interactions in IVR.

Hardware hiccups

The surveyed studies featured a range of sensorimotor-interaction enabling hardware, which could introduce another confound. Although it unclear how the hardware enabling sensorimotor input affects input, experience or learning, initial research has found that these devices are not neutral in their impact. For example, one study found that that the Leap Motion hand tracker is preferred by participants to the HTC wand when grabbing close objects due to its natural affordances; but also resulted in much slower performance at the selection of distance objects [83]. How we might deal with the impact of the affordances of the hardware are not clear, but studying, understanding and noting the impact of these at a system and experiment design stage could be beneficial.

An unconnected field

Finally, it was notable how few of these papers referred to each other. Among the papers, there were only four references to other studies found in this selection. Three of these [332][115][139] referred to one paper [50], while the final reference was to one of these three ([257] referencing [332]) (see Fig. 3.2). These references were all contained within studies investigating language learning; there were no cross-learning topic references.

This suggests that IVR sensorimotor learning has not yet established a central canon, which could bring challenges to researchers looking for similar works, understanding their outcomes and exploring and critiquing other research procedures and approaches.

3.1.6 Conclusion

This review found that sensorimotor-based IVR learning explorations are happening across a variety of subjects, backed predominantly by theories of embodied cognition. There are two two preeminent justifications for the research: (1) to provide empirical evidence for an existing theory of learning; and (2) to understand if IVR is a suitable replacement or improvement on physical-world teaching. A variety of measures for learning were also found, as well as a vast selection of experiential variables being measured.

From these results, a number of shortcomings for the field were summarised, which resulted in suggestions for further research for future sensorimotor-based IVR learning practitioners. These are:

(1) Be mindful of the potential confound of situation and context (situation as location-in-space, whereas context is more cognitive);
(2) Engage in more longitudinal or holistic topic learning, rather than singular interventions;
(3) Define measures of success, including beyond learning gain;
(4) Identify viable learning outcomes for interventions;
(5) Establish and engage sensorimotor-orientated embodiment questionnaires;
(6) Understand and discuss how the chosen IVR hardware might impact results;
(7) Engage with other research in the field, potentially across traditional discipline divisions

It is important to be mindful of these findings in the experiment design of studies in this thesis, and to understand and engage with how the experiments in this thesis have been designed and conducted in relation to the findings above. The results from this review were produced after the first experiment (Chapter 4.1), and changes informed by these results for both the second (Chapter 4.2) and third experiments (Chapter 4.3) include (1) the use of an abstract environment instead of a situated one; (3) exploring response time as a potential measure of learning gain; and (5) leveraging existing embodiment questionnaires and using sensorimotor experience-specific questions. The identification of viable learning outcomes (4) is discussed in the next section (Chapter 3.3).

The lack of longitudinal learning investigations (2) is discussed in detail in the future work section (Chapter 5.3.3); while the impact of chosen IVR hardware (5) and stronger engagement with wider research in the field (7) is presented in the background section (Chapter 2).

3.2 Conducting IVR experimental research remotely: surveying approaches and expert opinion

3.2.1 Introduction

Many IVR studies take place in laboratories with the co-presence of the researcher and the participant [168]. However, the COVID-19 pandemic highlighted the importance and perhaps necessity of understanding and deploying remote experimentation and/or recruitment methods within IVR research. As much of the research in this thesis was conducted during the COVID-19 pandemic, in order to conduct the research required for this thesis, a remote experimentation process was potentially the only viable method for data collection.

However, the IVR research community has been slow to embrace recruiting remote participants to take part in studies running outside of laboratories - a technique which has proven useful for non-IVR HCI, social and psychological research [249][235]. There is also limited literature about conducting remote IVR experimentation - although what reports exist suggest that the approach shows promise: data-collection is viable [304], results are similar to those found in-lab [222] even when the participants are unsupervised [131], and recruiting is possible [191]. Researchers have also suggested using existing communities for these technologies, such as customisable social IVR experiences, as combined platforms for recruitment and experimentation [271].

Due to the limited literature and the relatively new technology in this area, it was important to understand the approach in more detail before creating and deploying remote experiments. As the existing literature is limited, this research attempts to understand the benefits and limitations around conducting remote research by surveying active researchers who use or are investigating the use of IVR technologies for remote experimentation.

This section outlines the methodology and results from perhaps the first survey of IVR and AR researchers regarding remote IVR and AR research. The results are derived from 46 respondents answering 30 questions regarding existing research practice. It offers three core contributions: (1) its summarises existing research on conducting remote IVR and AR experiments; (2) it provides an overview of the status quo, showing that many of the concerns regarding remote IVR and AR are those also applicable to other remote studies; and that the unique aspects of remote IVR and AR research could offer more benefits than drawbacks; and (3) it sets out recommendations for advancing remote IVR and AR research, and outline important questions that should be answered to create an evidence-backed experimentation process.

Due to an overlap in some of the unique properties of spatial or immersive computing, both IVR and AR researchers are included in this research (see [260] for a summary of perspectives on the similarities and distinctions between VR and AR). Going forward, 'XR' is used as umbrella term for IVR, augmented reality (AR) and mixed reality (MR) [190].

3.2.2 Literature review

This literature review discusses relevant publications on XR research, remote research and remote XR research. The chapter is organised in three parts. First, it explores conventional XR experiments under 'normal' conditions (e.g. in laboratory andor directly supervised by the researcher). It then summarises existing literature on remote experiments in XR research. Finally, it reports the main findings from previous publications on remote data collection and experimentation.

Conventional XR experiments

Experiment types and fields of interest

According to Suh and Prophet's 2018 systematic literature review [310], XR experiments involving human participants can broadly be categorised into two groups: (1) studies about XR, and (2) studies about *using* XR. The first group focuses on the effects of XR system features on the user experience (e.g. if enhancing embodiment could affect presence outcomes [258]), whereas the second category examines how the use of an XR technology modifies a measurable user attribute (e.g. if leveraging XR embodiment could affect learning outcomes [256]). Across these categories there have been a variety of explorations on different subjects and from different academic fields. These include social psychological [27], including social facilitation–inhibition [129], conformity and social comparison [26], social identity [161]; neuroscience and neuropsychology [168], visual perception [341], multisensory integration [56], proxemics [275], spatial cognition [335], education and training [251], therapeutic applications [89], pain remediation [112], motor control [64], terror management [144] and media effects such as presence [13].

The theoretical approaches behind these studies are also disparate, including theories such as conceptual blending, cognitive load, constructive learning, experiential learning, flow, media richness, motivation, presence, situated cognition, the stimuli-organism-response framework and the technology acceptance model [310].

Data collection, approaches and techniques

According to Suh and Prophet's meta-analysis, the majority of XR research explorations have been experiments (69%) [310]. Other types of explorations include surveys (24%), interviews (15%) and case studies (9%). These approaches have been used both alone and in combination with each other. Data collection methods are predominantly quantitative (78%), although qualitative and mixed approaches are also used. Another systematic review of XR research (focused on higher education) [251] adds focus group discussion and observation as research methods, and presents two potential subcategories for experiments: mobile sensing and "interaction log in IVR app", in which the XR application logs the user's activities and the researcher uses the resulting log for analysis.

The types of data logging found in XR experiments are much the same as those listed in Weibel's exploration of physiological measures in non-immersive virtual reality [339], with studies using skin conductance [351], heart rate [75], blood pressure [124], as well as electroencephalogram (EEG) [3]. Built-in inertial sensors that are integral to providing an XR experience, such as head and hand position for IVR HMDs, have also been widely used for investigations, including posture assessment [34], head interaction tracking [353], gaze and loci of attention [247] and gesture recognition [152], while velocity change [337] has also been used in both IVR and AR interventions

Benefits of XR experiments

There are many suggested benefits to using XR technology as a research tool: it allows researchers to "control the mundane-realism trade-off" [5] and thus increase the extent to which an experiment is similar to situations encountered in everyday life without sacrificing experimental control [27]; to create *powerful sensory illusions within a controlled environment (particularly in IVR)*, such as illusions of self-motion and influence the proprioceptive sense [301]; *improve replication* [27] by making it easier to recreate entire experimental environments; and allow *representative samples*[27] to experience otherwise inaccessible environments, when paired with useful distribution and recruitment networks.

Challenges of XR experiments

Pan [233] explored some of the challenges facing experiments in virtual worlds, which continue to be relevant in immersive XR explorations. These include the challenge of ensuring the *experimental design* is relevant for each technology and subject area; ensuring a consistent feeling of self-embodiment to ensure engaged performance [162]; avoid *uncanny valley*, in which characters which look nearly-but-not-quite human are judged as uncanny and are aversive for participants [220]; *simulation sickness* and nausea during IVR experiences[221]; *cognitive load* [314] which may harm participation results through over-stimulation, particularly in IVR [306] [204]; *novelty effects* of new technology interfering with results [61] [78]; and *ethics*, especially where experiences in IVR could lead to changes in participants' behaviour and attitude in their real life [14] and create false memories [285].

Remote XR experiments

There has been little research into remote XR experimentation - experiments that takes place outside of a researcher-controlled setting - particularly for IVR and AR HMDs. This is distinct from field or in-the-wild research, which is research "that seeks to understand new technology interventions in everyday living" [266], and so is dependent on user context. These definitions are somewhat challenged in the context of remote IVR research, as for IVR, remote and field/in-the-wild are often the same setting, as the location where IVR is most used outside the lab is also where it is typically experienced (e.g. home users, playing at home [191]).

In terms of remote XR research outcomes, Mottelson and Hornbæk [222] directly compared in-lab and remote IVR experiment results. They found that while the differences in performance between the in-lab and remote study were substantial, there were no significant differences between effects of experimental conditions. Similarly, Huber and Gajos explored uncompensated and unsupervised remote IVR samples and were able to replicate key results from the original in-lab studies, although with smaller effect sizes [131]. Finally, Steed et al. showed that collecting data in the wild is feasible for virtual reality systems [304].

Ma et al. [191] is perhaps the first published research on recruiting remote participants for IVR research. The study, published in 2018, used the Amazon Mechanical Turk (AMT) crowdsourcing platform, and received 439 submissions over a 13-day period, of which 242 were eligible. The participant demographics did not differ significantly from previously reported demographics of AMT populations in terms of age, gender, and household income. The notable difference was that the IVR research had a higher percentage of U.S.-based workers compared to others. The study also provides insight into how remote XR studies take place: 98% of participants took part at home, in living rooms (24%), bedrooms (18%), and home offices (18%). Participants were typically alone (84%) or in the presence of one (14%) or two other people (2%). Participants reported having "enough space to walk around" (81%) or "run around (10%)". Only 6%

reported that their physical space would limit their movement.

While Ma et al's work is promising in terms of reaching a representative sample and the environment in which participants take part in experiments, it suggests a difficulty in recruiting participants with high-end IVR systems, which allow six-degrees of freedom (the ability to track user movement in real space) and leverage embodied controllers (e.g. Oculus Rift, HTC Vice). Only 18 (7%) of eligible responses had a high-end IVR system. A similar paucity of high-end IVR equipment was found by Mottelson and Hornbæk [222], in which 1.4% of crowdworkers had access to these devices (compared to 4.5% for low-end devices, and 83.4% for Android smartphones). This problem is compounded if we consider Steed et al's finding that only 15% of participants provide completed sets of data [304].

An alternative approach to recruiting participants is to create experiments inside existing communities of XR users, such as inside the widely-used VR Chat software [271]. This allows researchers to enter into existing communities of active users, rather than attempt to establish their own. However, there are significant limitations for building experiments on platforms not designed for experimentation, such as programming limitations, the ability to communicate with outside services for data storage, and the absence of bespoke hardware interfaces.

Remote data collection and experimentation

Validity, benefits, drawbacks and differences

Using networks for remote data collection from human participants has been proven valid in some case studies [108, 169]. In Gosling et al's comprehensive and well-cited study [108], internet-submitted samples were found to be diverse, generalise across presentation formats, were not adversely affected by non-serious or repeat respondents, and present results consistent with findings from in-lab methods. There is similar evidence for usability experiments, in which both the lab and remote tests captured similar information about the usability of websites [324].

That said, differences in results for lab and remote experiments are common [308, 37, 286]. The above website usability study also found that in-lab and remote experiments offered their own advantages and disadvantages in terms of the usability issues uncovered [324]. The factors that influence differences between in-lab and remote research are still being understood, but even beyond experiment design, there is evidence that even aspects such as the participant-perceived geographical distance between the participant and the data collection system influences outcomes [219].

Reips' [262] well-cited study outlined 18 advantages of remote experiments, including (1) easy access to a demographically and culturally diverse participant population, including participants from unique and previously inaccessible target populations; (2) bringing the experiment to the participant instead of the opposite; (3) high statistical power by enabling access to large samples; (4) the direct assessment of motivational confounding; and (5) cost savings of lab space, person-hours, equipment, and administration. He found seven disadvantages: (1) potential for multiple submissions, (2) lack of experimental control, (3) participant self-selection, (4) dropout, (5) technical variances, (6) limited interaction with participants and (7) technical limitations.

Supervised vs unsupervised

With the increasing availability of teleconferencing, it has become possible for researchers to be co-'tele'present and supervise remote experiments through scheduling webcam experiment sessions. This presents a distinction from the unsupervised internet studies discussed above, and brings its own opportunities and limitations.

Literature broadly suggests that unsupervised experiments provide suitable quality data collection [269, 120, 159]. A direct comparison between a supervised in-lab experiment and a large, unsupervised web-based experiment found that the benefits outweighed its potential costs [269]; while another found that a higher percentage of high-relevance responses came from unsupervised participants than supervised ones in a qualitative feedback setting [120]. There is also evidence that unsupervised participants react faster to tasks over the internet than those observed in the laboratory [159].

For longitudinal studies, research in healthcare has found no significant difference between task adherence rates between unsupervised and supervised groups [67]. However, one study noted that supervised studies had more effective outcomes [172].

Crowdworkers: Viable?

Remote data collection was theorised to bring easy access to participants, including diverse participants and large samples [262]. Researchers have found that recruiting crowdworkers, people who work on tasks distributed to them over the internet, allowed them access to a large participant pool[235], with enough diversity to facilitate cross-cultural and international research [39]. Research has found that crowdworkers were significantly more diverse than typical American college samples and more diverse than other internet recruitment methods [39], at an affordable rate [235][39]. This has allowed researchers a faster theory-to-experiment cycle [209].

Results from crowdworker-informed studies have been shown to reproduce existing results from historical in-lab studies [235] [39] [302], while a direct comparison between experiment groups of crowdworkers, social media-recruited participants and on-campus recruitment, found almost indistinguishable results [43].

Some distinctions between crowdworkers and in-lab have been discovered, however. Comparative experiments between crowdworkers and in-person studies have suggested slightly higher participant rejection rates [302], while participants have been shown to report shorter narratives than other groups of college students (both online and in-person) and use proportionally more negative emotion terms than college students reporting verbally to an experimenter [113].

Distinctions also exist within crowdworker recruitment sources. A study of AMT, CrowdFlower (CF) and Prolific Academic (ProA) found differences in response rate, attention-check question results, data quality, honesty, diversity and how successfully effects were reproduced [239].

Data quality is a common concern regarding crowdworkers [105]. However, attention-check questions used to screen out inattentive respondents or to increase the attention of respondents have been shown to be effective in increasing the quality of data collected [8], as have participant reputation scores [240].

A growing concern regarding crowdworkers is non-naivete, in which participants having some previous knowledge of the study or similar studies that might bias them in the experiment. Many workers report having taken part in common research paradigms [234], and there are concerns that if researchers continue to depend on this resource, the problem may expand. As such, further efforts are needed by researchers to identify and prevent non-naive participants from participating in their studies [40].

Summary

It is clear that remote methods have been usefully deployed for non-XR research, and seemingly bring benefits such as easier participant recruitment, reduced recruitment cost and broadened diversity, without introducing major biases. However, there is still a paucity of research regarding the extent to which remote XR research can and has been used to leverage the unique benefits of both XR (environmental control, sensory illusions, data collection, replication) and remote (participation, practicality, cost-savings) methods, as well as the potential impact of their combined limitations. Therefore a survey of XR researcher experiences and beliefs regarding remote XR research could help us understand how these apply practically at the current time, and understand the key areas for future developments in this field.

3.2.3 Methodology

Survey

The current practice of researchers was surveyed to understand researcher-perceived benefits and drawbacks of lab-based and remote XR research. A 30-item qualitative questionnaire was used that enquired about participants' existing lab-based and remote research practices; thoughts on future lab-based and remote research; and potential benefits and drawbacks for each area. The survey was circulated through relevant mailing lists (visionlist@visionscience.com, BCS-HCI@ jiscmail.ac.uk, chi-announcements@listserv.acm.org), to members of groups thinking of or currently running remote studies, and to members of universities' virtual and augmented reality groups found via search engines.

Responses were thematically analysed using an inductive approach based upon Braun and Clarke's six phases of analysis [31]. The coding and theme generation process was conducted twice by independent researchers; themes were then reviewed collaboratively to create the final categorisations.

Participants

There were 46 responses to the survey from 36 different (predominantly academic) institutions. Most responses came from researchers based in Europe and North America, but responses also came from Asia. The majority of participants were either PhD students (18) or lecturers, readers or professors (11) at universities. Other roles were academic/scientific researcher (5), masters student (5), corporate researcher (4) and undergraduate student (2). A diverse set of ages responded to the survey: 18-24 (5), 25-34 (22), 35-44 (11), 45+ (6), and gender skewed male (29) over female (16) or other (1).

3.2.4 Participant XR setup results

Participants were more likely to have previously ran in-lab studies (37) than remote studies (14). Twenty-seven participants noted that, because of the Covid-19 pandemic, they have considered conducting remote XR experiments. In the next six months, more researchers were planning to run remote studies (24) than lab-based (22).

Usefully for this thesis, participants predominantly categorised their research as IVR-only (28) over AR-only (5). Ten participants considered their research as both IVR and AR (and three did not provide an answer). This result is

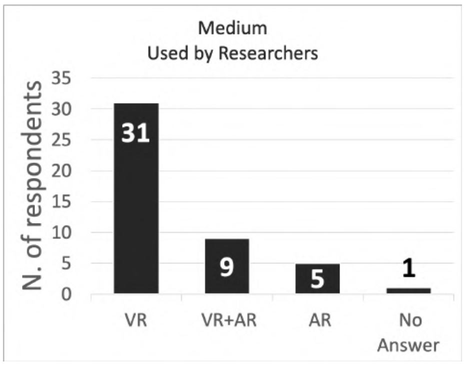


Figure 3.3: Type of XR medium explored by survey respondents.

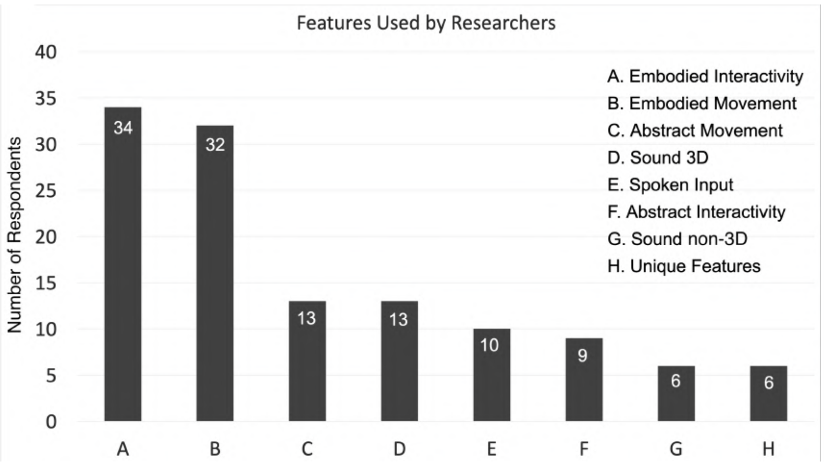


Figure 3.4: Features used by respondents in their user studies. (A) Embodied Interactivity: using embodied controller/camera-based movement. (B) Embodied Movement:using your body to move/'roomscale'. (C) Abstract Movement: using a gamepad or keyboard and mouse to move. (D) Sound 3D: binaural acoustics. E) Spoken Input. (F) Abstract Interactivity: using a gamepad or keyboard and mouse to interact. (G) Sound non-3D: mono/stereo audio. (H) Unique features: e.g. haptics, hand tracking, scent.

illustrated in Fig. 3.3. In terms of research hardware, the majority of IVR research leveraged HMD-based systems with six degrees of freedom (32), that tracks participants' movements inside the room, over three degrees of freedom (15) or CAVE systems (1). Nineteen researchers made use of embodied or gesture controllers, where the position of handheld controllers are tracked in the real world and their position virtualised. For AR, HMDs were the predominant medium (13) over smartphones (9), with some researchers (5) using both.

An array of supplementary technologies and sensors were also reported by 13 respondents, including gaming joysticks, haptic actuators, a custom haptic

Method	Summary
In-lab (vital)	Experiment requires features only feasible in-lab, e.g. bespoke hardware, unique data collection
In-lab (preferred)	Concerns about integrity of data collected remotely, high value on controlled setting
Remote (vital)	User's natural (in-the-wild) environment is important (e.g. Social VR, naturally experienced at home and online)
Remote (preferred)	Priority to get cross-cultural feedback or reach large number of participant; lab provides limited benefits

Table 3.2: Summary of XR Study Sub-types

glove, motion capture systems, e-textiles, eye-trackers, microphones, computer screens, Vive body trackers, brain-computer interfaces, EEG and electrocardiogram (ECG) devices, galvanic skin response sensors and hand-tracking cameras, as well as other spatial audio and hardware rigs.

The use of a variety of different off-the-shelf systems was also reported: Vive, Vive Pro, Vive Eye, Vive Index, Vive Pucks, Quest, Go, Rift, Rift S, DK2, Cardboard, Magic Leap One, Valve Knuckles, Hololens. Predominantly used devices are part of HTC Vive (25) and Oculus (23) family.

Respondents outlined numerous features of immersive hardware that they used in their research, visible in Fig. 3.4. The most prominent were embodiment aspects, including embodiment interactivity, in which a user's hand or body movements are reflected by a digital avatar (37) and embodiment movement (35), where participants can move in real space and that is recognised by the environment. Abstract movement (13), where a user controls an avatar via an abstracted interface (like a joystick) and abstract interactivity (8) were less popular. Spoken input was also used (10), as well as 3D sound (13) and non-3D sound (6). Scent was also noted (1) along with other unique features.

3.2.5 Thematic analysis results

In this section, the themes found in the survey study are presented and discussed. The key points of each theme are summarised in a table at the start each subsection. Some of these points were found across multiple themes as they touch various aspects of user-based XR research.

Key Point	Issue	Lab	Remote
Recruitment Scope	Sample size	Usually smaller numbers	Potential for larger number
Recruitment Scope	Sample balance	Might be easier to ensure balance	How to ensure balance? (e.g. who mostly owns XR equipment?)
Efficiency	Time	Requires setup time and organise participants	Potential less time especially if encapsulated and unsupervised
Precursor Requirements	Requisites	Pre-test and linguistic/culture comprehension conditions are ensured	Not clear how to verify conditions in remote studies

Table 3.3: Study Participants Key Points

Theme: Study Sub-types

The analysis suggests that in-lab and remote studies can be additionally distinguished by whether the setting type is vital or preferred (summarised in Table 3.2). Broadly, in-lab (vital) studies require experimental aspects only feasible in-lab, such as bespoke hardware or unique data collection processes; in-lab (preferred) studies could take place outside of labs, but prefer the lab-setting based upon heightened concerns regarding the integrity of data collected and place a high value on a controlled setting. Remote (vital) studies are required when a user's natural environment is prioritised, such as explorations into behaviour in Social VR software; and remote (preferred) studies are used when cross-cultural feedback or a large number of participants are needed, or if the benefits offered by an in-lab setting are not required.

Beyond these, another sub-type emerged as an important consideration for user studies: *supervised* or *unsupervised.* While less of an important distinction for in-lab studies (which are almost entirely supervised), participant responses considered both *unsupervised* 'encapsulated' studies, in which explanations, data collection and the study itself exist within the software or download process, and *supervised* studies, in which researchers schedule time with the remote participant to organise, run and/or monitor the study. These distinctions will be discussed in more detail throughout the analysis below, as the sub-types have a distinct impact on many of the feasibility issues relating to remote studies.

Theme: Study Participants

Recruitment scope

Twenty-nine respondents stated the well-known challenge of recruiting a satisfactory number of participants for lab-based studies. Issues were reported both with the scale of available participants, and the problem of convenience sampling and WEIRD - Western, educated, industrialized, rich and democratic societies - participants[119].

Participant recruitment was mentioned by 27 respondents as the area in which remote user studies could prove advantageous over labs. Remote studies could potentially provide easier recruitment (in terms of user friction: accessing to lab, arriving at the correct time), as well as removing geographic restrictions to the participant pool.

Removing the geographic restrictions also simplifies researchers' access to cross-cultural investigations (R23, R43). While cross-cultural lab-based research would require well-developed local recruitment networks, or partnerships with labs in target locations, remote user studies, and more specifically, systems built deliberately for remote studies, introduce cross-cultural scope at no additional overhead.

There are, however, common concerns over the limitations to these benefits due to the relatively small market size of XR technologies. The penetration of IVR HMD technology is currently limited, and it is possible that those who currently have access to these technologies will not be representative of the wider populations. Questions remain over who the IVR HMD owners are, if they exhibit notable differences from the general population, and if those differences are more impactful than those presented by existing convenience sampling.

Despite the belief that designing for remote participants will increase participant numbers, and therefore the power of studies, it seems unclear how researchers will reach HMD-owning audiences. Thirty respondents who have, or plan to, run remote XR studies have concerns about the infrastructure for recruiting participants remotely. Unlike other remote studies, the requirement for participants to own or have access to XR hardware greatly reduces the pool (around 5 million XR HMDs were sold in 2020 [317]). A major outstanding question is how researchers can access these potential participants, although some platforms for recruiting XR participants have emerged in the past few months such as XRDRN.org.

Nine respondents noted that remote XR experiments may encourage participation from previously under-represented groups, including introverts and those who cannot or do not wish to travel into labs to take part (e.g. people who struggle to leave their homes due to physical or mental health issues).

However, respondents with research-specific requirements also raised concerns that recruitment of specific subsets of participants could be more difficult remotely. For example, when recruiting for a medical study of those with age-related mobility issues, it is unlikely that there will be a large cohort with their own XR hardware.

Theme: Efficiency

Twenty-five respondents noted the potential for remote studies to take up less time, particularly if remote studies are encapsulated and unsupervised. They stated that this removes scheduling concerns for both the researcher and the participant, and allows experiments to occur concurrently, reducing the total researcher time needed or increasing the scale of experiment. However, there are concerns this benefit could be offset by increased dropouts for longitudinal studies, due to a less "close" relationship between research and participant (R17, R25).

Participant precursor requirements

One respondent noted they needed to run physiological precursor tests (i.e. visual acuity and stereo vision) that have no remote equivalent. Transitioning to remote research has meant this criteria must now be self-reported. Similarly, experiments have general expectations of linguistic and cultural comprehension, and opening research to a global scale might introduce distinctions from typically explored population. One respondent cautioned that further steps should be taken to ensure participants are able to engage at the intended level, as in-lab these could be filtered out by researcher intuition.

Data Collection

The overwhelming drawback of remote XR research, as reported by the majority respondents, was that of data collection. Excluding changes to participant recruitment, as mentioned above, the issues can broadly be categorised as: (1) bespoke hardware challenges, (2) monitoring/sensing challenges, and (3) data transmission and storage.

The use of bespoke hardware in any type of remote user study is a well-known issue, predominantly regarding the difficulty of managing and shipping bespoke technology to participants and ensuring it works in their test environments. In the context of XR technologies, 13 respondents voiced concerns about the complicated and temperamental system issues that could arise, particularly surrounding the already strenuous demands of PC-based IVR on consumer-level XR hardware, without additional overheads (e.g. recording multiple cameras).

Key Point	Lab	Remote
Hardware	Access custom and/or reliable hardware	Limited access to devices (e.g. EEG, ECG, computational power, etc.)
Data	Collection can be supervised, more detailed, real-time, more space for qualitative	Mostly unsupervised (less control), human expressions (e.g. facial) are generally lost, qualitative feedback is harder to collect
Behaviour	Likely more serious, richer (qualitative) data	Lack of detailed feedback, potentially less honest

Table 3.4: Data Collection Key Points

Four respondents felt it was unreasonable to ask remote participants to prepare multiple data-collection methods that may be typical in lab-studies, such as video recording and motion tracking. There were also concerns regarding the loss of informal, ad-hoc data collection (e.g. facial expressions, body language, casual conversations).

Finally, concerns were also raised regarding the efforts required to encapsulate all data capture into the XR experience, the effects this might have on data collection (for example, a recent study highlighted a difference on the variability of presence when participants recorded it from inside the IVR experience versus outside [284]), the reliability of transferring large amounts of data from participants, and how sensitive information (especially in the context of medical XR interventions) can securely be transferred and stored. This areas perhaps presents the biggest area for innovation for remote XR research, as it is reasonable to assume the academic community could create efficient, easy-to-use toolkits for remote data collection in XR environments which integrate to ethics-compliant data archives.

Many data collection methods were deemed infeasible for remote experimentation: EEG, ECG, eye/hand tracking, GSR, as well as body language and facial expressions. Five researchers noted adaptions they had been working on to overcome these, including using HMD orientation to replace eye tracking, and using built-in HMD microphones to record breaths instead of ECG monitoring to determine exertion, or using the HMD controllers to perform hand tracking.

Respondents also noted some behavioural concerns and changes for remote, unsupervised participants. These included a lack of participation in qualitative feedback (6 respondents); for one researcher (R20), participants were “encouraged to provide feedback but few took the initiative.” Another researcher (R31) stated “Debriefing is such a good space to collect unstructured interview data.

Key Point	Issue	Lab	Remote
Process & Guidance	Control	Full control over setup and participants	No control and guidance over participants
Process & Guidance	Participants	Rapport with researcher, welcoming, more serious, attentive	Different attitude, potential cheating
Environment	Setting	Can be distracting (e.g. outside noise) but generally more controlled	Might be distracting or overwhelming but likely more realistic/natural for participants
Hardware & software	Hardware	Access to custom devices, normal calibration process	No calibration (by researcher), potential for unknown errors, no custom tools
Hardware & software	Software	Allows for Wizard of Oz, adjust setting in real time	Issues harder to spot and influence results, longer development time
Research questions	Topics	Unchanged, if we go back to normal research conditions	Remote setup might influence research questions and topics
Cost	Expenditures	More time consuming, more expensive to run	Potentially cheaper but potentially more work for implementation

Table 3.5: Experiment Process Key Points

Users relax after the questionnaire/debriefing ... produc[ing] a ... meta-narrative where participants consider your questions and their experiences together”. The lack of supervision raised concerns regarding whether participants were being “truthful” in their responses, with one researcher (R41) stating that participants attempted to “game” their study in order to claim the participation compensation. However, others stated that unsupervised studies could reduce research bias arising from their perception of the participants’ appearance and mannerisms.

Theme: Experiment Processes

Process & Guidance

Many respondents were concerned that unsupervised participants may conduct the experiments incorrectly, or have incorrect assumptions, or misunderstand processes or target actions. Twenty-four respondents felt that guidance would be better provided (introduction, explanations, etc) in a lab setting that also allows ad-hoc guidance and real-time corrections.

There were also concerns over the mental state of participants: remote participants “may not take it seriously” or not focus (lack of motivation and engagement) or approach the study with a specific mood unknown to the researcher (R19, R30). Contrasting opinions suggested that participants may feel that the in-lab experience is “overly formal and uncomfortable” (R32).

Some respondents stated that remote experiments risk losing the “rapport” between researcher and participant, which might negatively influence the way a participant performs a remote study. However, one respondent stated that the transition to remote experimentation allowed them different, deeper, on-going connection with their participants. Their research was for a nIVR machine learning tool, and they found that moving away from in-person experimentation and to a remote workshop process encouraged the up-take of longitudinal community-building tools. The chosen communication method between researcher and user - Discord servers - became a place for unsupervised interaction between participants, and led to an on-going engagement with the research (R33). However it should be considered that any “rapport” between participant and researcher might introduce bias.

Environment

Concern was raised around participants’ environments, and their potential varying unsuitability for remote experimentation, compared with controlled laboratory settings. For example, one respondent (R20) stated: “one user reported walking into their whiteboard multiple times, causing low presence scores.” The concern is particularly strong for unsupervised remote experiments, as distractions could enter into the experiment and affect data without the researcher being aware.

This concern was not universal, however. Four respondents noted that their laboratories space was far from distraction free, and even suggested that a remote space could prove freer of interruptions than the space available to them in their research setting; while others stated that researchers should be mindful that the laboratory itself is an artificial space, far more so than where people will typically use their IVR setups - in their homes. Five respondents highlighted how XR research could benefit from being deployed in “the participants’ own environment”.

The immediate environment of the user was also raised as a concern for IVR experiment design: the choice of being able to move freely in an open space in a laboratory against a more adaptive solution for the unknown variables of participants’ home environments.

Respondents noted that supporting the different IVR setups to access a larger remote audience would also prove more labour-intensive, and would introduce more variables compared with the continuity of the tech stack available in-lab. With remote experiments, and more so for encapsulated unsupervised ones, 10 respondents believe there will be more time spent in developing the system.

Hardware and software

A concern regarding remote experiments, particularly unsupervised, is that calibration processes are harder to verify (R30). This could either cause participants to unknowingly have faulty experiences, and therefore report faulty data; or it will increase time taken to verify user experiences are correct. Unknown errors can effect data integrity or participant behaviour. Respondents noted that this type of remote error are often much more difficult and labour-intensive to fix compared with in-lab. This issue is compounded by individual computer systems introducing other confounding factors (for both bug-fixing and data collection) such as frame-rates, graphic fidelity, tracking quality and even resolution can vary dramatically.

Five respondents reflected that overcoming these issues could lead to more robust research plans, as well as better development and end-product software to overcome problems listed. This encapsulation could also lead to easier opportunities for reproducability, as well as the ability for researchers to share working versions of the experiment with other researchers, instead of just the results. It could also help with the versioning of experiments, allowing researchers to build new research on-top of previous experiment software.

Four respondents were aware these advantages are coupled with longer development times. The increased remote development requirements could also be limiting for researchers who face constrained development resources, particularly those outside of computer science departments. This is compounded by the fact that the infrastructure for recruiting remote XR participants, data capture, data storage and bug fixing is not particularly developed. Once these are established, however, respondents felt these might make for a higher overall data quality compared with the current laboratory-based status quo, due to more time spent creating automated recording processes, and not relying on researcher judgement. There are also arguments that the additional development time is offset by the potential increase in participants and, if unsupervised, the reduction in experiment supervision requirements.

Six respondents that use specific hardware in their research, noted that it was currently difficult to measure physiological information in a reliable way, and included hand tracking in this. However, we are aware that some consumer IVR hardware (Oculus Quest) allows hand-tracking, and so there is an additional question of whether researchers are being fully supported in knowing what technologies are available to them.

To alleviate issues with reaching participants, two respondents wrote about potentially sending equipment to participants. The limitations of this were noted as hardware having gone missing (which had happened, R35), and participants being unable to use equipment on their own (which had not happened yet).

Research questions

Five respondents noted that their research questions changed or could change depending on whether they were aiming for a laboratory or remote settings. For example, one respondent (R31) suggested that “instead of the relationship of the physical body to virtual space, I’d just assess the actions in virtual space”. Others explored the potentiality of having access to many different system setups, for example, now being able to easily ask questions like “are there any systematic differences in cybersickness incidence across different HMDs?”. (R39)

Nine respondents speculated that remote research has potential for increasing longitudinal engagement, due to lower barriers to entry for researcher (room booking, time) and participant (no commute), and that rare or geographically based phenomena could be cheaply studied using remote research; as providing those communities access to IVR may be cheaper than relocating a researcher to them.

Costs

Eight respondents noted the potential of remote experimentation for reducing some of the cost overheads for running experiments. Laboratories have important costs that are higher than remote studies: lab maintenance, hardware maintenance, staff maintenance. Without these, costs per participant are lower (and for unsupervised studies, almost nil). As experiment space availability was also noted as a concern for laboratory-based experiments, this seems a potentially under-explored area of benefit, provided remote participant recruitment is adequate.

Theme: Health & Safety

The leading benefit given for remote user studies was that of health and safety, citing shared HMDs and controllers as a potential vector for Covid-19 transmission, as well as more general issues such as air quality in enclosed lab spaces.

Key Point	Summary
Protocols	Missing standard protocols (to work safely with participants in-lab)
Equipment	Sanitizing of in-lab equipment and spaces
Remote	Concerns for remote participants (e.g. accidents during a user study)
Real-Time Aid	Not available for remote participants (e.g. motion sickness)

Table 3.6: Health and Safety Key Points

Concerns were raised for both viral transmission between participants, and between participant and the researcher. This concern has also increased administration overheads, with 6 respondents stating it could be more time consuming to prepare the lab and organise the studies or using new contract-tracing methods for lab users.

However, respondents also raised concerns about additional safety implications for remote participants. The controlled lab environment is setup to run the study, whereas remote participants are using a general-purpose space. Additionally, for health and mental health studies, in-lab allows for researcher to provide support, especially with distressing materials. Finally, IVR environment design has a direct impact on the level of simulator sickness invoked in participants. There were questions about the responsibility of researchers to be present to aid participants who could be made to feel unwell from a system they build.

Theme: Ethics

Three ethics concerns were reported by respondents: encouraging risky behaviours, responsibility for actions in XR and data privacy. An example of this might be the ethical implications of paying participants, and therefore incentivising them, to take part in what could be considered a high-risk behaviour: entering an enclosed space with a stranger and wearing an IVR HMD.

Respondent (R30) raised the question of liability for participants who are injured in their homes while taking part in an XR research project. The embodied nature of XR interventions - and most respondents used this embodiment in their studies - could put participants at a greater risk of harming themselves than with other mediums.

Finally, while cross-cultural recruitment was seen as a potential boom for remote research, questions were raised about ethics and data storage and protection rules when participants are distributed across different countries, each

Key Point	Summary
Suspensions	No user studies at the moment
Facilities	Sanitizing of equipment and spaces
Recruitment	Harder/impossible to recruit in-lab participants
Exclusion	Bias and high risk participants

Table 3.7: Covid-19 Implications

with different data storage laws and guidelines. Although not limited to XR, due to the limited number of IVR users, and the disproportionate distribution of their sales, it seems the majority of remote IVR participants will originate from North America, and ethics clarification from non-US-based universities are needed.

3.2.6 Covid-19 Implications

While COVID-19 has impacted most studies around the world, the dependence on shared hardware for XR research, especially HMDs, has led to many implications reported by respondents. These concerns are particularly related to COVID-19, and therefore be reduced as the pandemic is resolved. However, as it is currently unclear when the pandemic will end, we felt it was useful to discuss them in a dedicated section.

Most respondents noted that COVID-19 had caused a suspension of studies and that they were unclear how long the suspension would last for, resulting in an overall drop in the number of studies being conducted, with 30 respondents stating it will change the research they conduct (e.g. moving to online surveys). The continuation of lab studies was eventually expected, but with added sanitizing steps. However for many, it was unclear what steps they should take in order to make XR equipment sharing safe. These concerns extended beyond the XR hardware to general facility suitability, including room airflow and official protocols which may vary for each country and/or institution.

Five respondents also had concerns about participants. There were worries that lab-based recruitment would be slow to recover, as participants may be put off taking part in experiments because of the potential virtual transfer vectors. Similarly, respondents were concerned about being responsible for participants, and putting them in a position in which their is a chance they could be exposed to the virus.

There was also concerns around COVID-19 and exclusion, as researchers who are at high risk of COVID-19 or those who are in close contact with high risk populations, would now have to self-exclude from lab-based studies. This might introduce a participant selection bias towards those willing to attend a small room and sharing equipment,

It should be noted that not all labs are facing the same problems - some of our respondents had continued lab-based experimentation during this period, with COVID-19 measures ensuring that participants wore face masks during studies. This was considered a drawback as combined with an HMD, it covered the participant's entire face and was cumbersome. These measures are also known not to be 100% protective.

3.2.7 Discussion

The previous section presented the themes found in the analysis. Some of these presented common characteristics and some issues were reported in multiple themes. These results have been summarised in the next section, highlighting the key points and suggesting important questions for future research.

Recruitment and participants

As with non-XR experiments, researchers are interested in the potential benefits of remote research for increasing the amount, diversity and segmentation of participants compared with in-lab studies. However, with many respondents reporting that it has been difficult to recruit XR participants, it seems there is a gap between potential and practice. The unanswered question is how to build a pool of participants that is large and diverse enough to accommodate various XR research questions and topics, given that there are few high-end HMDs circulating in the crowdworker community [191][222]. So far, we have found three potential solutions for participant recruitment, although each requires further study:

(1) Establish a dedicated XR crowdworker community. However, concerns of non-naivety[234], which are already levied at the much larger non-XR crowdworker participant pools, would surely be increased. We would also have to understand if the early version of this community would be WEIRD[119] and non-representative, especially given the cost barrier to entry for HMDs.

(2) Leverage existing consumer XR communities on the internet, such as the large discussion forums on Reddit. These should increase in size further as they shift from early-adopter to general consumer communities. However, these communities may also have issues with representation.

(3) Establish hardware-lending schemes to enable access to a broader base of participants [305]. However, the cost of entry and risk of these schemes may make them untenable for smaller XR research communities.

It is also not clear, beyond HMD penetration, what the additional obstacles are that XR poses for online recruitment. Technical challenges (e.g. XR applications needing to run on various devices, on different computers, requiring additional setup beyond simple software installation) and unintuitive experiment procedures (e.g. download X, do an online survey at Y, run X, record Z) for participants are notable distinct issues for remote XR research. It is also unclear if the use of XR technology has an impact on what motivates participants to take part in remote studies, an area of study that has many theoretical approaches even in the non-XR area[160].

Data collection

Respondents feel that many types of physiological data collection are not feasible with either XR or non-XR remote research. For remote XR research, there are unique concerns over video and qualitative data collection as using XR technologies can make it (technically) difficult to reliably video or record the activity, as well as moving participants' loci of attention away from the camera or obscuring it behind an HMD. However, the hardware involved in creating XR experiences provides a variety of methods to gather data, such as body position, head nodding, breath-monitoring, hand tracking, HMD angle instead of eye tracking. These can be used to explore research topics that are often monitored via other types of physiological, video or qualitative data, such as attention, motivation, engagement, enjoyment, exertion or focus of attention. It would be useful for XR researchers to build an understanding of what the technologies that are built into XR hardware can tell us about participant experiences, so as to allow us to know the data collection affordances and opportunities of XR hardware.

That said, the infrastructure for collecting and storing this (mass) of XR data remotely is currently not fully implemented, and we are not aware of any end-to-end standardised framework. However, work is being done to simplify the data collection step for XR experiments build in Unity [34]. There are also opportunities to further develop web-based XR technologies that could send and store data on remote servers easily. There are also ethical concerns, as respondents were unclear on guidance regarding data collection from participants located in other nations, particularly when they should be paid. This includes how the data is collected, where it should be stored, and how can be manipulated.

Health, safety and Covid-19

During this research, as a result of the COVID-19 pandemic, many laboratories were considered unsafe for running user studies. Although some respondents reported being able to continue working in-lab, the limitation meant it was feasible for many researchers to run user studies under normal conditions. The main concern during that period is the lack of standardised protocols to ensure safety of researchers and participants while running user studies and the issue with the ethics protocols of the research institutions, especially during potential future epidemics or pandemics. For XR research, it is unclear how to adequately sanitize equipment and tools, as well as how to maintain physical distancing. There are also concerns about the comfort of participants if they are required to wear masks alongside HMDs. Finally, respondents reported concerns about a potential long-term fall in user motivation to take part in such experiments, when HMDs are a notable infection vector. There are distinctly different safety and ethics concerns around remote XR experiments, including the research responsibility for not harming participants (e.g. ensuring environments are safe for the movements, and not inducing simulator sickness), which, while also true of in-lab experiments, are considered a greater challenge when a participant is not co-located with the researcher.

Mediated impact

A concern for respondents was that remote settings introduce additional uncontrolled variables that need to be considered by researchers, such as potential unknown distractions, trust in participants and their motivation, and issues with remote environmental spaces. However, previous research shows that most HMD-wearing remote participants engage in space well-known to them (the home) and predominantly when they are alone [191], which could alleviate some of the environmental space and distraction concerns. Further research into how a home environment could impact XR studies is needed, and the creation of well-defined protocols to alleviate uncontrolled influences remote XR results. Beyond this, we also need to understand any impact that remote experiments may have on results compared with in-lab experiences, especially if we are to be able to reliably contrast lab and remote research. Previous research for non-XR experiments suggest that distinctions between lab and remote settings exist [37][286] [308], but it has been theorised that the impact might be less for XR experiments, as you “take the experimental environment with you” [27].

‘Encapsulated’ experiments: the ideal?

Respondents stated that creating remote XR experiments might encourage better software development and experimental processes. If experiments are able to be deployed as all-in-one experience and data collection bundles that can run unsupervised, the time-saving implications for researchers (and participants) are huge, especially when paired with the potential increase in participants. This type of ‘encapsulated experiment’ can also improve replication and transparency, as theorised by Blascovich [27], and allow for versioning of experiments, in which researchers can build on perfect replicas of other’s experimental environments and processes. Finally, due to the similar nature of XR hardware, data logging techniques could easily be shared between system designers or standardised; something we have seen with the creation of the Unity Experiment Framework [34].

However, there are some limitations to this approach. It is likely it will require additional development time from the researchers, especially as a comprehensive experiment framework is established. In addition, there are data collection limitations for remote XR studies, as discussed in previous sections. It is also interesting to consider how encapsulation might work for AR investigations, as the environment will only partially be controlled by the designer.

Perhaps the potential for remote XR experiments lies in understanding the data collection affordances of the hardware; collectively building frameworks to ease the collection of this data; and to design research questions that maximise their use; all inside encapsulated experiences. This might be a mindset shift for researchers, who according to our survey, are predominantly lab-orientated.

3.2.8 Limitations

The goal with this research was to provide an overall insight into the perception of advantages and drawbacks of remote experimentation within the XR researcher community. However, this approach means that insights from sub-communities may not have been found. For example, we had no responses from researchers involved in topics such as vulnerable populations. Further investigation into sub-communities is needed to uncover potential insights for those areas.

3.2.9 Conclusion and Recommendations

It is clear from that respondents believe that remote XR research has the potential to be a useful research approach. However, it currently suffers from numerous limitations regarding data collection, system development and a lack of clarity around participant recruitment. Analysis of our survey results and literature around remote and remote XR research suggest that, to better understand the boundaries of remote XR experimentation, researchers need answers to the following questions:

(1) Who are the potential remote XR participants, and are they representative?

(2) How can we access a large pool of remote XR participants?

(3) To what extent do remote XR studies affect results compared with in-lab?

(4) What are the built-in XR data collection affordances of XR hardware, and what can they help us study?

(5) How can we lower the barriers to creating encapsulated experiment software, to maximise the potential of remote XR research?

There is also an opportunity to reconceptualise approaches to XR and remote research. XR experiments, as it stands, are predominantly used to study a participant's experience with an XR system, in an artificial but controlled setting (laboratory) using external data collection methods (surveys, cameras, etc.).

However, if researchers consider XR devices primarily as data-collection hardware with set properties, it is possible to work backwards to understand what research questions are suitable with the existing data collection afforded by XR hardware. Additionally, there is potential to reconceptualise, for suitable applications, the home as a natural research location and move away from the laboratory as the default location for user studies. This is a potentially unique opportunity for XR compared with non-XR studies as, for many investigations, the XR experiment takes the environment with it.

These findings suggested that, in response to the COVID-19 pandemic, conducting this thesis' (second and third) experiments remotely should not pose any major barriers to the accuracy or validity of the research. Indeed, by leveraging an encapsulated study design (2, 5), the number of participants was notably increased from the the in-lab study used for the first experiment, although the approach provided a far less representative population in terms of gender (1).

3.3 Current practice for designing an experiment measuring language learning

There are many considerations when designing an experiment to test language learning. It seems there is limited literature with dedicated research design guidance on conducting language learning experiments. Therefore an inductive analysis of a sample of language learning research interventions was used to inform the experimental design used in this thesis. The literature included in this search was not identified systematically, but based upon papers encountered that featured experimental language learning gain tests.

Three notable factors in learning gain experiment design were outlined: word types (Chapter 3.3.1); target language (Chapter 3.3.2); and (Chapter 3.3.3). Common approaches to these factors are discussed below, as well as the relevant design outcomes for the experiments used in this thesis.

3.3.1 Word types

Many investigations into sensorimotor activity and language learning have used verbs as their learning topic [7][71][106]. This result is unsurprising, given the consistent links found between verbs and sensorimotor activity; and between the regions of the brain that regulate sensorimotor activity and language [154][195]. Therefore a corpus including verbs was the obvious choice for the experiments in this thesis.

Less prominent in literature was learning where nouns were the target learning and object of sensorimotor interactions, although investigations do exist [194]. There is lower prevalence of evidence of a detectable impact on noun memorisation and sensorimotor activity, and so nouns were considered a secondary area for investigation in this thesis, given that the predominant exploration of this research was to find evidence that types of sensorimotor activity in IVR can cause distinct cognitive outcomes. That said, the first experiment includes both verbs and nouns to give some level of comparative understanding.

At the time of writing, research covering sensorimotor explorations for adverbs, adjectives or other linguistic factors were not found.

3.3.2 Target language

The choice of the target language in the experiment involves a few important decisions. The first is whether a living language or an artificial one is used. Some researchers have started conducting language learning experiments using artificial corpuses [87] (such as Vimmi, as used in [194]) of target words, instead of living languages. This is in order to avoid associations with participants' native or foreign languages. It was difficult to find resources testing the viability of these corpuses and their effects on participants, however, and as such it was felt safer to use vocabulary that could more closely reflect actual language learning goals of participants. It was a concern that learning an artificial language may impact motivation, a factor and potential confound that this thesis was interested in.

The concern raised above, regarding associations with the words in the learning experience and participants' native or foreign languages, is common in language acquisition literature. Therefore when selecting a language for testing, it is important to chose one that minimises these associations and to experimentally test how this prior knowledge may influence results.

A potential useful measure for this is 'linguistic distance', the extent to which languages differ from each other. Although no concrete measure of linguistic distance has been agreed upon, one popular measure is based on the difficulty Americans have learning other languages [54]. This rated Japanese and Korean as the most-distant languages for US speakers. This helped inform the choice of Japanese learning in the experiments in this theses. The other factors were existing language knowledge of the researcher and previous successful use of Japanese in a well-known verb with gesture learning intervention [155].

Some measure of control or understanding about participants' pre-knowledge of the target language is typical of learning gain investigations. Popular approaches include asking participants to self-exclude; excluding participants with high prior knowledge, detected via pre-testing, excluding data from participants when they already have pre-existing knowledge of that word, and comparing the mean pre-existing knowledge of participants of each experiment condition to ensure there is no significant discrepancy.

An artificial corpus allows for researchers to easily control for factors that may have an influence on memory, such as word length or number of syllable. A living language is constrained to the actual meanings assigned to those objects or actions, which may provide disproportionately challenging words in some situations. Therefore while using a living language, extra care must be taken in the vocabulary selection process around word length and initial or final phoneme similarities.

A notable issue with using Japanese as a corpus is that the applied language includes many *garaigo*, or loan words, from other languages, which could help learners intuit their meaning. Therefore care had to be taken to ensure *garaigo* words were excluded.

3.3.3 Testing approaches

A majority of language learning gain experiment research use quasi-experimental methods, measuring the learning gain difference of an intervention between two or more treatment groups. There appearance to be no particular dominance between between-subject and within-subject design.

An ancillary metric that is occasionally used as either a proxy for learning gain, or to quantify the depth of encoding, is answer response time. The conception is that a faster response time equates to better encoding of the language knowledge.

Learning retention is often measured one week after the initial intervention, although retention studies have lasted months [194][87] and years [225].

Part of the testing approach with the most variability between research designs is the number of words to be encoded in the session. These vary greatly, with examples ranging from six to thirty. With no concrete guidance, experiments in this thesis relied on piloting the first and second experiments and iterating based on qualitative feedback to achieve a learning size that did not feel too long for participants.

3.3.4 Conclusion

This inductive research highlights some established approaches regarding language learning gain experiments: the quasi-experimental process, the monitoring of words learned as the primary metric of learning (and response time as secondary) and the common length of retention periods being one week. There is a notable gap in literature providing guidance on, or engaging with, the recommended number of words learned in an experimental session, but generally around 10 - 20 words were used. These findings were used to guide the design of the experiments presented in this thesis.

Chapter 4

Evaluation

This chapter presents the rationale, methodologies, results and discussions of three experiments conducted to investigate if and how HBSI in IVR can influence cognition and memorisation, via a second language learning process. All learning discussed in this chapter specifically refers to verb or noun second language learning.

A summary of these experiments can be found in Table 4.1, including the number of participants, learning gain measurements and other measured factors.

The first experiment was designed to evidence if HSBI had an impact on learning outcome in IVR; and whether that was a positive one. As prior research suggests this would be the case [332], it was also designed to understand if any learning benefit involving HSBI could be attributed to a direct relationship between the sensorimotor interaction and learning, or whether it was mediated through another factors, such as presence, motivation or cognitive load.

This first experiment was also designed to serve as a pilot for future studies, testing the virtual environment and learning process and gathering feedback from participants, and also testing the experiment process.

The second experiment was designed to build upon the findings of the first experiment, attempting to further unpack how HBSI in IVR worked in relation to virtual object interactions. Literature from outside of IVR suggested that there are distinctions in learning outcomes when language knowledge has been encoded by either actions or gestures [334]. This experiment was designed to understand if similar distinctions might persist in IVR, which would present some understanding of how users cognitively contextualise the interactions they make in IVR, and how they considered virutal object manipulations.

The third experiment was designed to build upon the findings of the second experiment by exploring feedback-from-HBSI, rather than the HBSI itself. This was in order to understand more about HBSI in IVR design for learning activities, and also understand which cognitive theories of embodied interaction may be most accurate to describe the observed phenomenon.

The experiments were conducted sequentially, and a summary of these experiments and their explorations is presented below:

- Exp. 1) Investigating the possibility and mechanisms of HBSI benefits for language learning in IVR, answering the following questions:
  - Can learning occur in IVR?
  - Does HBSI-enabled learning outperform non-HBSI-enabled learning?
  - Can we attribute the learning benefit to a direct relationship between sensorimotor interaction and learning, or is it mediated through other factors, such as presence, motivation or cognitive load?
  - How did participants qualitatively respond to the sensorimotor-enabled and non-sensorimotor-enabled conditions?
- Exp. 2) Investigating if there is a distinction in learning outcome between (i) HBSI using virtual objects (defined as actions); and (ii) HBSI that does not interact directly with virtual objects (defined as gestures)
- Exp. 3) Investigating if there a distinction in learning outcome between (i) HBSI in which virtual objects present aural and visual interactive feedback; and (ii) sensorimotor interaction in which virtual objects do not provides aural or visual interactive feedback

The experimentation processes for all three experiments were approved by the QMUL Research Ethics Committee (QMERC), and participants were provided information based upon the QMERC Participant Information Sheet and agreed to the QMERC Consent Form. The experiments were deemed low risk and were reviewed by the Research Ethics team.

Investigation	Participants	Learning Measure	Other factors
HBSI vs non-sensorimotor interaction	24/24	Learning gain	Presence (single-factor), Player Learning Experience/Motivation (MEEGA+), Cognitive Load (single-factor)
Actions vs gestures	48/35	Learning gain, response time	Presence (single-factor), Presence (four-factor: general spatial, involvement, realism), Motivation (IMI), Embodiment
High interaction feedback vs low feedback	54/35	Learning gain, response time	Presence (four-factor: general spatial, involvement, realism), Motivation (IMI), Realism, Interactivity and Environmental impact

Table 4.1: Summary of the three experiments, including number of participants (shown as two numbers: those who completed the first session; and those who completed an addition retention test one week later), learning measurement(s) and other measured factors

4.1 Comparing sensorimotor and non-sensorimotor interactions

4.1.1 Introduction

This section presents the rationale, methodology, results and discussion regarding an experiment that compares participant learning between two IVR interaction conditions: HBSI and non-HBSI-enabled.

This comparison is important for the investigation of two questions that are fundamental for this thesis: a) can language learning occur in IVR; and b) does sensorimotor-enabled interaction impact participant learning gain?

This experiment also begins to explore the cognitive mechanisms through which the sensorimotor-enhanced memorisation occurs. It does this by monitoring factors that are considered related to IVR learning outcomes: feelings of presence, motivation and cognitive load. It attempts to determine if sensorimotor-enabled interaction has a mediating effect on one or more of these, which then led to enhanced learning gain; or whether learning gain improvements are directly related to the sensorimotor activity without a mediation effect.

To do these two things, an IVR environment for learning Japanese language

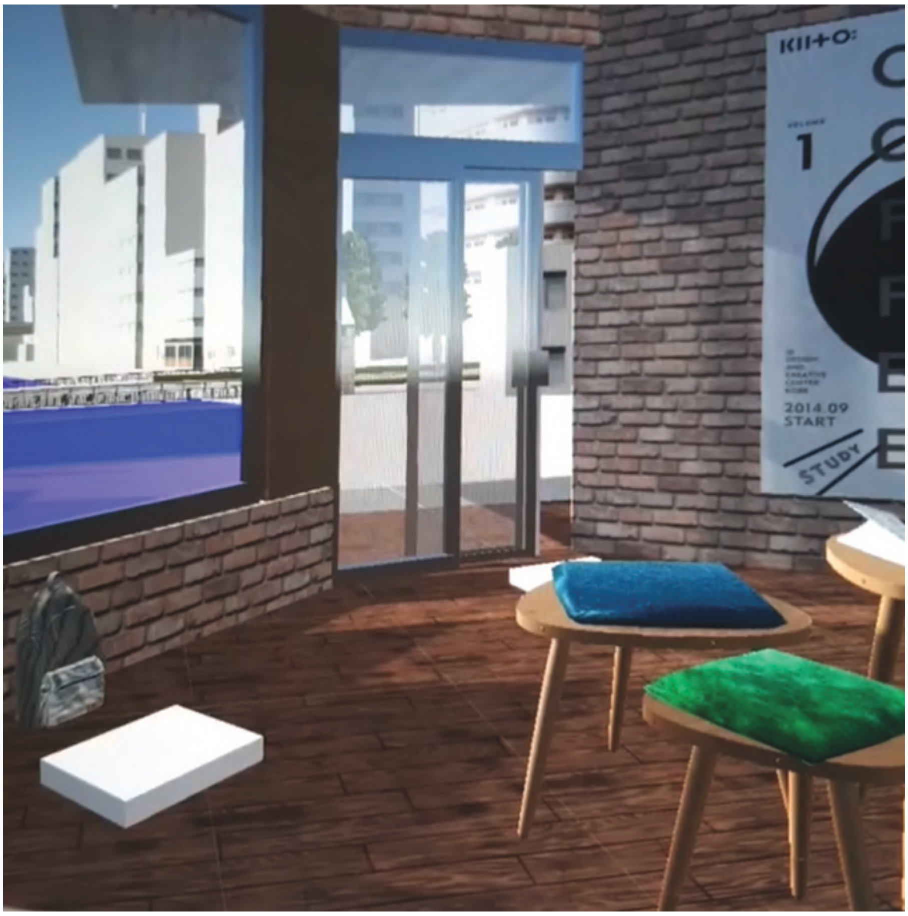


Figure 4.1: Image of the virtual reality environment used in this experiment. Picture shows an attempted realistic environment, with the “open the door” (middle) and “move the bag” (left) interaction points marked by white cuboids

words was created and participants were randomly split between two interaction conditions: sensorimotor-enabled and non-sensorimotor-enabled (see Fig. 4.2.

Learning gain was calculated by comparing pre- and post-exposure learning outcomes, both overall and with a distinction between words describing the participants’ actions (verbs) and words for the object used to conduct the action (nouns). This method of determining if learning occurred was based upon the inductive investigation of other language learning experiments mentioned in Chapter 3.3.

To understand what cognitive processes the different sensorimotor interaction conditions may have affected, cognitive load, motivation and presence were monitored as co-variables. To detect if results were influenced by poor system design, system usability was surveyed.

Additionally, participants were asked to provide semi-structured qualitat-

Hypothesis	**Outcome**
H1. Language learning occurs when learning inside IVR	Not rejected
H2. Sensorimotor-enabled interaction produces stronger learning outcomes than non-sensorimotor-enabled learning	Not rejected
H3. Sensorimotor interaction had no impact on cognitive load, presence or motivation	Not rejected
H4. Higher motivation scores correlate with higher learning gains	Partially rejected
H5. Higher presences scores correlate with higher learning gains	Not rejected
H6. The relationship between presence and learning gains is mediated by motivation	Rejected
H7. The relationship between sensorimotor-enabled interactivity and learning gains is mediated by motivation	Rejected
H8. The relationship between sensorimotor-enabled interactivity and learning gains is mediated by presence	Rejected

Table 4.2: Summary table of hypotheses and results

ive feedback about their experience in order to understand perceptions regarding learning a language in IVR, insights into the experience and experimental design, and to see if any unprompted feedback arose concerning sensorimotor interaction.

4.1.2 Experiment

This experiment was designed to explore two groups of hypotheses. The first group (H1 - H3) are concerned with comparing the two interaction conditions, sensorimotor-enable interaction and non-sensorimotor-enabled interaction:

- H1. Language learning occurs when learning inside IVR
- H2. Sensorimotor-enabled interaction produces stronger learning outcomes than non-sensorimotor-enabled learning
- H3. Sensorimotor-enabled interaction had no significant impact on cognitive load, presence or motivation

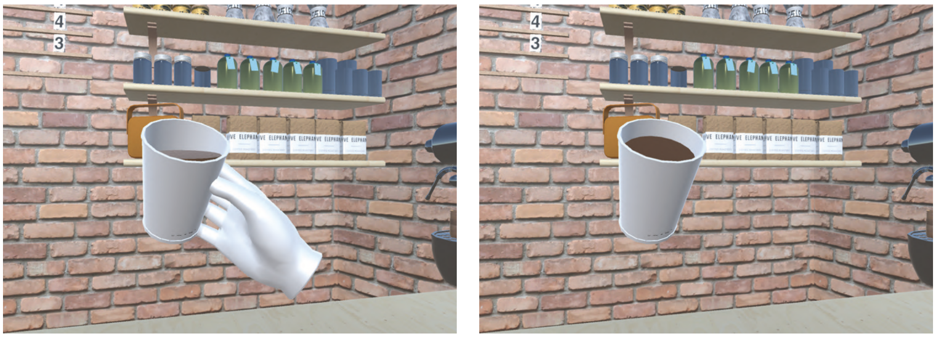

Sensorimotor-enabled interaction Non-sensorimotor-enabled-interaction

Figure 4.2: Showing HBSI in the sensorimotor-enabled condition, where participants have to manipulate 3D objects; and in the non-sensorimotor-enabled condition, in which the 3D objects are animated and participants do not manipulate them. This is the “take the drink” interaction point (see Table 4.3 for a full list of words and their interaction activities)

The second group of hypotheses are derived from literature regarding factors impacting IVR learning, such motivation, presence and embodiment, and attempt to understand if any effect on learning outcomes form sensorimotor-enabled interaction is mediated by one of these factors. These hypotheses are (also depicted in Fig. 4.3):

- H4. Higher motivation scores correlate with higher learning gains
- H5. Higher presences scores correlate with higher learning gains
- H6. The relationship between presence and learning gains is mediated by motivation
- H7. The relationship between sensorimotor-enabled interactivity and learning gains is mediated by motivation
- H8. The relationship between sensorimotor-enabled interactivity and learning gains is mediated by presence

Each participant was assigned to either an sensorimotor-enabled or non-sensorimotor-enabled group. They were then presented with 10 interaction areas inside an IVR coffee shop environment (see Fig. 4.1). Each interaction area contained an object and a related action. When a participant navigated to an interaction area, a voice-over introduced the object and explained the related action in both English and Japanese (e.g. “This is a jug. You can pour it. Pour in Japanese is sosogu; Say sosogu and pour the jug”; and after completing the action: “Jug is heiji. Say heiji”).

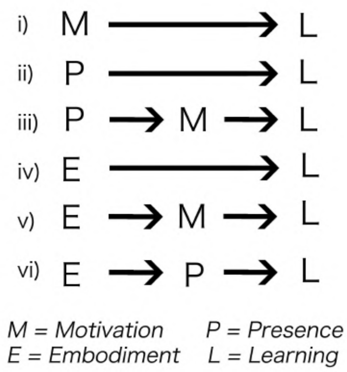


Figure 4.3: Approaches to relationships between motivation, presence, embodiment and learning. Examples iii, v and vi depict mediated relationships

Depending on their assigned group, when interacting with the object, the participant was asked to either:

- Speak the object and action words, and complete an accompanying action by grabbing and moving the related object using their sensorimotor-enabled controllers
- Speak the object and action words, and observe the related object move through an action without any input from the participant

Participants were introduced to each interaction area in sequence, then given 10 minutes to freely explore the environment and attempt to learn the words. For each interaction area, participants were first given the verb prompt: "This is a jug. You can pour it. Pour in Japanese is sosogu; Say sosogu and pour the jug". Participants could say the verb and perform the action concurrently or consecutively. After completing the action and saying the word, they were then prompted to say the name of the noun they performed the action with: "Jug is heiji. Say heiji".

Each participant only experienced one of the above conditions (between-subject design). A researcher was co-present and observed to ensure the participant entered the correct actions and spoken input. For incorrect actions or spoken inputs, the researcher could push a button to make the system re-prompt participants until they correctly performed the spoken utterance or action. When a correct spoken input (or a spoken and action input for the sensorimotor-enabled condition) was completed, the interaction area was marked as completed, the objects were reset to their original positions, and participants could repeat or visit other interaction points. Participants could leave or enter any interaction area at any point. During the learning process, the count of participants entering and exiting interaction areas, the time spent in each area and the success and failure rates of action or spoken production, were not recorded.

Memorisation: recognition, not recall

The memorisation tested in this experiment is recognition, rather than recall. In terms of learning, research suggests that recall is more effective for learning than recognition [99]; and can present a more accurate understanding of learner knowledge, given that the probability of guessing the correct response is higher in recognition than in recall testing [186]. However, recognition (e.g. multiple choice answer selection) is a widely-used measure of learning. There were also concerns about introducing subjectivity into the test marking process if assessing spoken recall was required, and regarding the ability of participants to be able to recall enough words, after a single intervention, to make for a useful study. Therefore, recognition was chosen as the type of memorisation for the experiment.

Why include spoken production?

This learning process uses spoken production as well as sensorimotor inputs. This was to ensure that the sensorimotor condition was tested while using other IVR-relevant modalities and naturalist learning approaches. Applied Linguistic research suggests that actively speaking second language words while learning causes increased word retention. This is known as the production effect [197], and plays an important role in modern second-language tuition. The positive impact of the production effect on computer-aided language learning (CALL) has been demonstrated experimentally, with automatic-speech recognition systems being some of the most effective types of computer-aided language learning tools [101]. Recent experiments combining gesture and spoken production have also demonstrated positive learning outcomes [155]. However, these experiments did not take place inside IVR, and whether and how these benefits transfer is yet to be investigated. By including spoken input here, the experiment tests the hypotheses in a situation similar to how IVR learning software could be designed, and also eliminates a confound in which some participants may naturally say the word aloud as part of their process, while some may not.

There is also evidence that multi-modal learning in IVR can cause cognitive load issues that harm information retention [280] [328] [204]. Therefore if spoken production was not included, this experiment may not uncover results pertaining to cognitive load that would exist in the wild.

Participants

Twenty-four uncompensated participants (15 male, 7 female) took part in the experiment. They were asked to self-report their knowledge of the target language (Japanese) and were pre-tested for their knowledge of the words used in the experiment. Participants were aged in ranges 30-39 (12), 20-29 (8) and 40-59 (4). No participants demonstrated an extensive knowledge of the target learning words during the pre-test (M = .13 (out of 20); SD = .44) nor self-rated their ability as anything above "basic phrases". Most participants were fluent in more than one language, but we did not find a significant difference in learning outcome between mono-lingual and multi-lingual participants ($t(22)$ = -.84, p = .20; mono-lingual: M = 6.17, SD = 3.18 ; multi-lingual: M = 7.83, SD = 4.25). Twenty-two participants were educated to post-graduate level or above. A visual inspection suggested there was not enough variance in answers related to interest levels in Japanese, Japan, virtual reality and coffee shops to prove useful for further analysis.

Corpus

Participants were tested on their knowledge of 10 noun/verb pairs (20 words). The included words were predominantly determined by considerations regarding the target sensorimotor interactions that participants were to enact to learn the words. The considerations included choosing sensorimotor interactions that (a) were likely familiar to participants, and so were not novel or required training to achieve; (b) used objects that needed to be manipulated to complete the action, to allow for both noun and verb study; (c) were not overly small or subtle, nor were potentially difficult to enact in IVR (i.e. sew a button); and (d) had a relatively unique sensorimotor movement compared with other actions in the set.

Further criteria included the technical complexity of both implementing the interaction and detecting the participant's successful completion of the interaction; and ensuring all choices were congruent with the setting (cafe) and with each other (using a cloth to wipe, rather than using a balloon).

Japanese gairaigo words (words imported from other languages, such as 'koohii' to mean 'coffee') were specifically avoided to reduce the chance of participants inferring a meaning. The list of included word pairs can be found in Table 4.3.

Environment

Learning took place in an IVR coffee shop environment created in Unity (2018.1.8), presented in a 'realistic' (non-stylised) style to provide a situated context for

Verb	**Noun**	**User Activity**
Take	Drink	Pick up the drink model from the counter
Pay	Money	Pick up the money model and put it in the payment location
Pour	Milk	Pick up the milk jug model and tilt it until milk pours out
Stir	Spoon	Pick up the spoon model and rotate it inside a cup model until a swirling animation appears
Cover	Lid	Pick up the lid model and place it on top of a cup model to cover the open top
Put	Tea	Pick up the black tea model and put it on a tray model
Eat	Cake	Pick up the cake model and bring it to the mouth until the model is replaced with a bitten one
Wipe	Napkin	Pick up the napkin model and wipe the surface so the peanut models are removed
Move	Bag	Pick up the bag model and move it to a different location
Open	Door	Grab the door handle and pull it along a sliding access until the door is opened

Table 4.3: Word pairs used and brief description of the activity

memorising words related to a coffee shop. The environment was explorable via a head-mounted display and sensorimotor-enabled controllers (the Oculus Rift S and Touch controllers).

Navigation could be done by moving around the real space, using the thumbsticks, or a combination of both. They could also rotate in the physical world, or by using the thumbstick to 'snap' turn 30 degrees at a time, in either direction. The movement speed was dependent on how far the thumbsticks were pressed, but subjectively may be described as slow, in order to reduce the risk of locomotion-induced simulator sickness. Participants could only pick-up objects by moving their hand so that the hand graphic was touching or embedded in the virtual object, and pressing the 'Grip' button on their controllers. The objects did not snap into a fixed position in the hand.

Evaluation

Participants' knowledge of the Japanese content was measured in three tests: one administered before their exposure to the environment (pre-test); one immediately after (post-test), and one seven days later (week-test). Participants performed the same test each time, listening to a Japanese word and typing the English (or another) language translation if they knew the meaning. The week-test was conducted remotely via the internet and not in controlled conditions. Participants were not given feedback when submitting answers. The maximum score was 20, and participants existing knowledge (i.e. their individual correct answers from the pre-test) were removed from their analysis to ensure only acquired knowledge was included in the results, which were analysed as a normalised result. Only three participants had any prior knowledge of the words, to very low levels. The pre-test results for correct answers for each group were $M = .01$, $SD = .03$ (sensorimotor-enabled) and $M = 0$, $SD = 0$ (non-sensorimotor-enabled).

After using the system, participants were asked to complete a MEEGA+ questionnaire [243] to provide insight on their motivation. MEEGA+ is designed to survey the quality of educational games. A reduced form of the "player experience" metric was used to provide motivational insight specific to a game-based learning experience. This involved removing the 'usability' category as well as questions regarding social interaction and a specific question regarding task monotonous, given that this was not an educational game but a research experiment. This left questions regarding confidence in learning outcome, challenge, satisfaction, fun, focused attention, relevance, and belief it would encourage the achievement of their short-term learning and learning goals. Participants were asked to rate statements across a five-point scale, from "strongly disagree", to "disagree", to "neither disagree nor agree", to "agree", to "strongly agree", with the final result a number between -2 (strongly unmotivating) and 2 (motivating).

Participants were also asked to self-report their cognitive load on a single-item, 9-point Likert scale as defined by Paas [231], which asked participants to rate their perceived amount of mental effort from "very, very low mental effort (-4)", to "very, very high mental effort (4)".

Finally, participants were asked to self-report the level of presence felt inside the environment on Slater's single-item, 6-point Likert scale [299], which asked participants to rate the extent to which they experienced a sense of "being there" inside the virutal environmental, ranging from "not at all really there (1)" to "totally there (6)". Asking participants for their subjective evaluation of presence experienced is considered the most direct way of presence assessment [135].

Copies of these questionnaires can be found in Appendix A. The results from these surveys are presented both as recorded and normalised, with the minimum and maximum possible results transformed to 0 and 1.

Learning Preferences

Participants were asked to complete the VARK learning preference questionnaire [86] to allow us to determine if learning preference would have an impact on results. However, there was too much homogeneity in the results to allow for segmentation analysis related to different learning preferences and so this was excluded from the analysis.

Analysis

In order to answer the questions (H1) can learning occur in IVR (as opposed to no learning); and (H2) does sensorimotor-enabled interaction improve learning outcomes over non-sensorimotor-enabled interaction, a one-tailed independent t-test on the learning gain from pre-test to post-test of the two interaction conditions. A one-tailed test was used as the expectation was an increased learning gain in the sensorimotor-enabled condition, as is found in non-IVR explorations.

Previous studies have suggested a distinction between the impact of sensorimotor-enabled interactivity on immediate post-exposure learning and that of longer-term retention, with learning that is sensorimotor-enabled improving retention but not immediate learning [332]. Thus immediate post-test and one-week later post-test conditions were analysed separately.

To understand if the sensorimotor interaction condition had an impact on cognitive load, presence and motivation (H3), a t-test was used for cognitive load; a Mann-Whitney U-test for presence (as data was not normally distributed); and a U-test for motivation (as motivation data did not meet the requirement of homogeneity).

In order to explore if higher motivation correlates with learning gain (H4); if higher presence correlates with learning gain (H5); and if motivation correlates with presence (H6); Pearson's r was used.

For exploring if the relationship between sensorimotor interaction and learning gains is mediated by motivation (H7) or presence (H8), a correlation matrix was used followed by multiple linear regression to understand potential contributions of motivation, presence and sensorimotor interaction method to learning gain scores. Baron and Kenny's mediation test [16] was used to deduce the mediation effects of presence on sensorimotor interaction; and motivation on sensorimotor interaction.

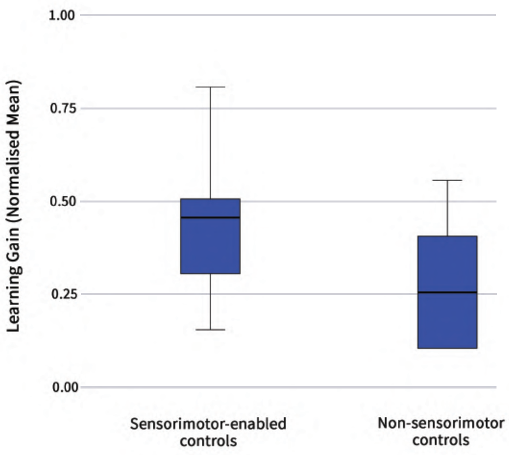


Figure 4.4: Showing difference in sensorimotor-enabled interactivity and non-sensorimotor-enabled interactivity gain

4.1.3 Results

H1, H2: Language learning occurs; sensorimotor-enabled interaction impact on learning gain

Results	N	Normalised Mean
Sensorimotor: Post-test	14	0.44 ±0.21
Non-sensorimotor: Post-test	10	0.28 ±0.16
Sensorimotor: Week-test	14	0.29 ±0.20
Non-sensorimotor: Week-test	10	0.18 ±0.11

Table 4.4: Table of normalised learning gain results (mean) and standard deviation from tests immediately after the session (post-test) and one week later (week-test)

Immediate learning gain

An independent-samples t-test was conducted to compare post-test learning gains in sensorimotor-enabled interaction and non-sensorimotor-enabled interaction conditions. For sensorimotor-enabled interaction ($M = 0.44$, $SD = 0.21$), memorisation was significantly higher than for the non-sensorimotor condition ($M = 0.28$, $SD = 0.16$), with a large effect size ($t(22) = 2.03$, $p = .027$, $g = .88$) (see Fig. 4.4 and Table. 4.4). This suggests that sensorimotor-enabled interaction had a large, meaningful benefit to immediate retention over the non-sensorimotor condition.

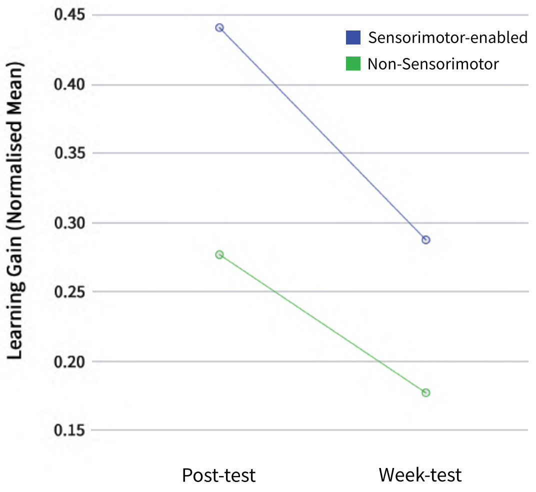


Figure 4.5: Showing the learning gain difference between sensorimotor conditions, and the drop in retention between the immediate test (post-test) and the test one week later (week-test)

When considering results for verbs and nouns separately, for verbs, there is a significant distinction between interaction conditions ($t(22) = 2.06$, $p = .026$, $g = .47$), but not for nouns $t(22) = 1.47$, $p = .078$, $g = .66$). This could suggest that the benefits of sensorimotor-enabled interactions are limited to verbs, or that nouns require additional sensitivity beyond that which the experiment provided.

The results for each condition were normally distributed, and met the requirements of homogeneity of variance.

Retention/one-week learning gain

Learning gain after one week for sensorimotor-enabled ($M = 0.28$, $SD = 0.20$) and non-sensorimotor-enabled ($M = 0.18$, $SD = 0.11$) conditions showed no significant difference ($t = 1.53$, $p = .07$, $g = 0.66$). Therefore while sensorimotor-enabled interaction promoted better immediate memorisation, this benefit was not detected for retention (see Fig. 4.5 for a depiction of immediate and one-week results by interaction condition).

The results for each condition were normally distributed, and met the requirements of homogeneity of variance.

Data validity

One participant's learning gain results were quite high compared with others, but they were not considered a significant outlier according to Grubbs' test, and thus included.

H3: Sensorimotor interaction type and cognitive load, presence and motivation

There was no significant difference between cognitive load scores for sensorimotor-enabled interaction ($M = .71$, normalised $= 0.088$) or non-sensorimotor-enabled interaction ($M = .50$, normalised $= 0.062$) conditions ($t(22) = .37$, $p = .36$).

There was no significant difference between the presence scores for sensorimotor-enabled interaction ($M = 3.29$, normalised $= 0.046$) or non-sensorimotor-enabled interaction ($M = 3.00$, normalised $= 0.40$) conditions ($U = 57$, $p = .23$) in a Mann-Whitney U-test.

There was no significant difference the motivation scores for sensorimotor-enabled interaction ($M = 0.71$, normalised $= 0.68$) or non-sensorimotor-enabled interaction ($M = 0.49$, normalised $= 0.62$) conditions ($U = 49$, $p = .12$).

H4: Motivation and learning gain

There was no significant evidence of a correlation between motivation and immediate learning gain ($r(22) = .29$, $p = .82$). However, there was a significant correlation between motivation and one-week learning gains, with a weak-to-moderate positive correlation ($r(22) = .35$, $p = .049$). See Fig. 4.6 for a plot of the individual data points.

H5: Presence and learning gain

There was a significant moderate correlation between presence with immediate learning gains ($r(22) = .41$, $p = .040$), and a significant, but weaker, correlation for one-week learning gains ($r(22) = .35$, $p = .045$. See Fig. 4.6 for a plot of the individual data points.

H6: Presence mediated by motivation

There was no significant evidence of a correlation between presence and motivation ($r(22) = .19$, $p = .16$).

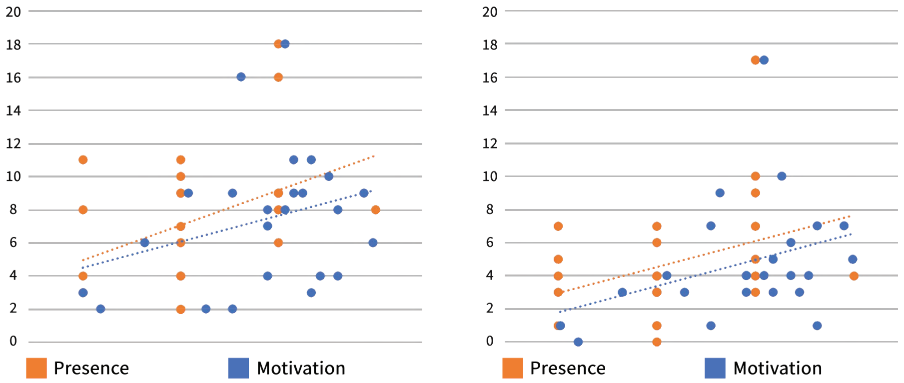
A summary of these can be seen in the correlation coefficient matrix in Table 4.5.

H7, H8: Sensorimotor-enable interaction and learning gain, mediated by motivation or presence

Immediate learning gain

	Immediate	One-week	Motivation	Presence	Sensori.
Motivation	0.29	0.35*	1		
Presence	0.41*	0.35*	0.19	1	
Sensorimotor	0.40*	0.31	0.31	0.18	1

Table 4.5: Correlation coefficient matrix between learning gains (immediate and one-week later), motivation, presence and sensorimotor interaction. * denotes significance



Immediate

One week

Figure 4.6: Graph showing the correlation between presence, motivation and test score; Y-axis showing test score, X-axis showing standardised presence and motivation results. A significant relationship was found between presence and immediate learning; and presence and one-week learning; and motivation and one-week learning. The results show that participants who felt more presence or motivation also achieved higher learning gain

A multiple linear regression showed that motivation, presence and sensorimotor interaction were not significant factors for impacting learning gain when all three were included in a model. Using a backward step-wise multiple linear regression, we found that presence ($Beta = .355$, $t(21) = 1.89$, p ¡ .05) and sensorimotor interaction ($Beta = .355$, $t(21) = 1.78$, p ¡ .05) explained a significant amount of the variance in immediate learning gain, and explained 19% of the learning outcome variance ($R2 = .192$).

Tests showed that multicollinearity was not a concern.

As sensorimotor interaction was not found to have a significant relationship with presence ($t(22) = 0.84$, $p = .21$), there is no evidence of a mediation effect of presence on sensorimotor interaction (according to Baron and Kenny's mediation test).

As motivation was not a significant regression of learning outcome, there was no evidence of a mediation effect of motivation on sensorimotor interaction (according to Baron and Kenny's mediation test).

One-week learning gain

There was evidence for relationships between presence and one-week learning gain, and motivation and one-week learning gain (see Table 4.5). Both showed weak-to-moderate significant correlations.

Multiple linear regression showed that motivation, presence and sensorimotor interaction predictors were not significant factors for impacting one-week learning gain when all three were included in the model ($R2 = .124$). Using backwards step-wise regression, we found that only presence ($Beta = .354$, $t(22) = 1.77$, p ¡ .05) explained a significant amount of the variance in the one-week learning gain.

As sensorimotor interaction was not found to be a significant regressor of one-week learning gain, there was no evidence that presence or motivation mediated sensorimotor interaction (according to Baron and Kenny's mediation test).

4.1.4 Discussion

H1, H2: Learning happens, and can be enhanced by sensorimotor-enabled interaction

The results show that learning occurred in the IVR, which does not reject H1. More interestingly, they also showed that sensorimotor-enabled interaction provides significant and notable immediate learning benefits over non-sensorimotor-enabled interaction, which does not reject H2. This means that in IVR, whether an experience uses sensorimotor-enabled interaction or not can have an impact on learning gain.

The results replicate findings in the physical world, in which the sensorimotor-enabled condition provides learning advantages [193] over the non-sensorimotor-enabled condition. This could suggest that sensorimotor-enabled learning benefits in the physical world may be continue to exist in the same way in virtual space.

However, a significant difference was only found for verb memorisation, and not for noun memorisation. One explanation for this could be, uncontroversially, that the learning topic plays an important role in how sensorimotor-enabled interaction affects the learning gain. As verbs refer to the actual sensorimotor activity performed, rather than the subject of the action (as the noun objects are), then this distinction could have a meaningful difference.

While the sensorimotor interaction type may have a more detectable effect on verb learning than on nouns, this does not mean that it is effect-less on nouns. Although the results for noun learning were non-significant, the mean learning gain for the sensorimotor-enabled condition was higher than for the non-enabled condition. This may suggest that the effect was harder to observe on nouns; one that this experiment's relatively small sample size was unable to pick up.

The order the words were presented in the experience may also have played a part: verbs were always presented to participants first, with nouns coming afterwards. Reflecting on the experiment process, visual observation of participants noted that some participants would first do the verb and noun, but on their second visit to an interaction area, they would only do the verb part of the interaction. This meant that for some users there was a potentially greater exposure to verb learning than noun learning.

The verb-first order was chosen for two reasons: (a) it was theorised that verbs had a stronger chance of showing a learning difference when encoded with sensorimotor activity, and as the study was designed to find if an effect could be observed in IVR, it seemed sensible to put the word type with the higher chance at the start of the learning interaction; (b) there was a concern that participants would, if not given an immediate sensorimotor interaction instruction, lose focus on the task.

Given the above, if a study was designed to specifically compare the sensorimotor interaction learning distinction for verb and noun learning, then it would be crucial to vary the order in which verbs and nouns were presented.

There was also a lack of significance in difference in retention between conditions. While the mean was higher for the sensorimotor-enabled condition, the lower learning gain after one week for both conditions meant that a greater sensitivity (i.e. more participants) could have been beneficial in detecting a distinction. Alternatively, it may be that sensorimotor-enabled interaction only has an impact on immediate learning gain. This would be an interesting result, as there are few examples of learning interventions that create a stronger immediate gain with no long-term benefits. This would be distinct from evidence found in the physical world, and could be a unique property of IVR-sensorimotor-enabled encoding. However, a previous IVR language learning experiment only found learning gain in the long-term and none in the short term [332], which refutes this idea. More evidence would certainly be needed to support this theory.

H3: Sensorimotor interaction had no impact on cognitive load, presence or motivation

The original plan was to monitor cognitive load to serve as a potential explanation if a non-significant learning difference between conditions was found. Previous research suggests that IVR, especially IVR with many interaction modalities, can harm learning through increased cognitive load demand [280]. The results presented here do not support the hypothesis that leveraging sensorimotor-enabled interactions increase cognitive load. Nor do they show that sensorimotor-enabled interaction reduces cognitive load, as proposed by Steed et al. [306]. The recorded difference in mean cognitive load for the sensorimotor-enabled condition (M = .71, normalised = 0.088) and the non-sensorimotor-enabled interaction (M = .50, normalised = 0.062) was both very small and not significant ($t(22) = .37$, $p = .36$).

However, there was also no correlation between reported cognitive load and participants' learning scores. The relationship between cognitive load and learning outcomes is widely evidence, and therefore this result could suggest that either measure of cognitive load used here was not sufficiently sensitive, or that the entire learning experience did reach a cognitively-harmful 'load' threshold to negatively impact learning.

There was no evidence for an impact of sensorimotor interaction type on self-reported levels of presence or motivation, both of which are often hypothesised. This could indicate that there is no relationship between HBSI and either presence or motivation.

However, while no evidence of a relationship between sensorimotor interaction and presence or motivation was found, it could also be that sensorimotor interaction is simply a more subtle contributor to presence or motivation then wearing a HMD, and therefore creates too small of an effect to be detected in this experiment, in which IVR was novel to most participants. Similarly, it is likely that, as the entire IVR experience was quite motivating for participants, and therefore any contribution of sensorimotor interaction to self-reported motivation could have been lost in this already highly motivating experience. Running the experiment with participants familiar with IVR might help remove some of these confounding factors.

Although more research is needed to uncover a stronger conclusion without the limitation noted above, the results do not reject H3.

H4: Motivation and learning gain

The results showed a mixed relationship between motivation and learning gain. Motivation did not significantly correlate with immediate learning gain, but was significant and showed a correlation with one-week learning gain. This suggests that either motivation is only an impactful contributor to forming long-term memorisations, or that the process was not sufficient to correctly understand the relationship between motivation and immediate learning gains. This does partially reject H4.

This result could be due to the choice to analyse "player experience", which focuses the motivation metric on the participants' experiences with the learning system itself, rather than wider motivation for the learning subject. A robust future exploration would likely include both the system-level motivation and a wider examination of participant motivations.

Another explanation could be that all participants were generally motivated to a similar level, with 22 out of 24 participants reporting positive motivation scores. This would be a potentially useful outcome for future studies concerned with investigating links between variables inside IVR, motivation and learning outcome, as it suggests that there are limitations in analysing motivation scores in an already highly-motivating experience (i.e. IVR).

H5: Presence and learning gain

The relationship between presence and learning outcome was shown to be significant and moderate for both immediate and one-week learning results, which does not reject H5. As the presence, motivation, learning paradigm was rejected, this result suggests that there is something implicit and important about presence itself that contributes to learning, and it is not simply a causal factor for motivation and its effects. This result reinforces many existing perceptions around presence as both a key affordance of IVR, and one that can have notable cognitive impacts and learning benefits.

H6: Presence, motivation, learning gain

The weak correlation and lack of a significant relationship between presence and motivation shows no evidence of presence enhancing motivation. As presence has a significant correlation with learning gain, these outcomes present no evidence that motivation has a complete mediation effect on presence, rejecting H6. This result is caveated by the limitations of the motivation result above.

H7, H8: Sensorimotor mediating factors

The results did not evidence mediating factors on sensorimotor interaction for either the immediate learning gains or the one-week learning gains, rejecting H7 and H8. Combined with the result that sensorimotor-enabled interaction was found to have a significant impact on learning gains, it seems clear that there is no complete mediating effect present (although it should be noted that there could still be partial mediating effects that are not evidenced in these results).

As no mediating effect was found, this result provides evidence for the view that sensorimotor interaction can have a direct impact on cognitive outcomes (in this case, evidenced by immediate verb learning), and the view that sensorimotor interaction is the second profound affordance of IVR.

This result also supports an embodied cognitive perspective of learning, in which the sensorimotor activity is directly related to cognitive, and does not create stronger memorisation by enhancing another factor linked to memorisation, such as motivation or presence.

4.1.5 Limitations

There are notable limitations to this study. The first is with the data collection methods for presence and motivation. The method for measuring presence was not comprehensive, as, although the one-question presence survey used here has been widely used and validated, and self-reporting is considered the most direct method for rating presence, a more thorough approach may have employed more in-depth measurements, such as measuring component aspects of 'presence', like (in one presence model) plausibility and place illusion. There are some concerns regarding whether questionnaires alone are suitable for establishing an accurate presence result [294]. Similarly, the measure of cognitive load would also be more robust if paired with additional physiological measures, such as electroencephalography [151], heart rate [232], or electrodermal activity [291].

The metric for motivation used in these results is defined by the player's experience of the learning system, and not their overall sense of motivation. The metric is generated from the participants' feelings of confidence, challenge, satisfaction, fun, focus and self-perceived relevance to their learning goals, which only provides participant-to-system motivation. Perhaps a better metric would examine intrinsic motivation for acquiring the target learning language, or for engaging in language learning or learning generally. The motivation scores reported by participants were also overwhelming positive, which limited the variance of the motivation factor. This high level of motivation may be explained by the novelty of IVR for the participants, and could be alleviated by using participants more experienced with IVR systems.

The environment was also designed to maximise the physicality of the learning, with grabbable nouns and verbs as the target learning acquisitions. Therefore caution should be used in trying to extrapolate these results for more abstract language concepts, such as adjectives, and for other learning subjects. And even for language acquisition, a longitudinal study would be more advantageous over a single-session learning intervention [132].

There was also a further distinction between the conditions beyond HBSI versus non-HBSI. This is that the HBSI condition involved the participant first grabbing an object and then doing an action with it. This HBSI could therefore be considered two distinct HSBIs in one: first grabbing, then doing. Contextualised from this perspective, it is important to understand if the 'grabbing', rather than the ‘action’, was a key contributor to learning enhancements in the HBSI condition. The evidence in this experiment is not strong enough to argue either way, and therefore this question would be worth exploring in further work interested in this area. However, in terms of practical applications, while the ‘grab’ from an action condition could be removed, it might be a rather unnatural experience for users; as often actions can not take place without first grabbing the item. There, it seems reasonable to consider the grabbing as part of the natural action movement.

We should also avoid extrapolating these results to language learning generally: the environment and its memorisation objective are non-natural, and focus on word, rather than language, acquisition. How some of the research’s outcomes – such as the benefits of sensorimotor-enabled action for verb acquisition – might contribute to other important aspects of second language acquisition, such as communicative competence; or for other learning subjects; is still unclear and not covered in this work.

4.1.6 Qualitative learnings

Participants were also surveyed regarding their experiences with the system, in order to understand perceptions regarding learning a language in IVR and if any unprompted feedback arose concerning sensorimotor interaction.

They were asked the following three free-form survey questions:

- Can you list three strong aspects of the system?
- Can you list three weak aspects of the system?
- Do you have any other comments?

A reflexive thematic analysis was conducted, with most feedback concerned with either the environment and the sense of spatial presence, interactivity and sensorimotor interaction, and general IVR feedback.

Environment and spatial presence

The IVR environment was broadly praised both for its realism (respondent #2, #5, #12) and the sense of “immersion” (#3, #8, #18) it offered.

Four respondents (#2, #15, #22, #23) praised the learning experience of target words in the contextually-relevant environment presented by IVR, with feedback like “it created a environment like reality, help me learn in situation”, “[the environment makes it] more effective than reading books” and “the environment and game definitely helped me connect the words and objects”. One respondent also (#16) noted that the the “spacial element of the game (elements position within the room) helps in associating a word with a physical position”.

However, the IVR soundscape had mixed feedback. Three participants stated that the music added to the overall sense of immersion in the environment (#3, #8, #18), however, four felt that it was distracting for their learning process (#3, #5, #16, #17).

Some participants also struggled with the fidelity of the environment, suggesting that there was too much detail or environmental stimulus (#5, #7, #8). Some participants reported that they would prefer “maybe fewer design details, since they made me pay attention to the space rather than the instructions”; “too many things in the world ... is distracting precisely because of how immersive it is”; and “I find my attention drifting as I wanted to move and interact with the world in a novel way”. One participant noted that the IVR environment was “distracting”, with “too much ’mental energy’ to move around and do actions that is subtracted from focusing/learning the words”.

Other feedback included that individual locations in the cafe were too similar for their optimal learning (#3) and two participants stated that target learning objects, particularly tea and coffee, were too similar (#12, #15).

Sensorimotor interaction

Many participants in the sensorimotor-enabled interaction condition said it was “fun” (#1, #6, #9 #10, #11) or that it “made the repetition quite fun” (#9), which suggests motivational or engagement benefits from engaging in sensorimotor-enabled interactions. Two participants mentioned that interacting with the objects in a natural way was “rewarding” (#6, #10), while another (#19) stated that “virtually touching objects makes you connect with the concepts you are trying to learn”.

IVR

Participants provided mixed feedback regarding topics related to IVR generally. Regarding navigating the virtual environment, two participants (#1, #21) stated the movement was easy and using a joystick to navigate was a positive aspect. However, ten participants made negative comments about location, including suggesting it caused feelings of simulator sickness (#4, #5, #9), it was difficult (#6, #10, #22), and that movement distracted from learning (#5, #16, #19, #20). One participant suggested that it would be good if objects could be interacted with from a distance, without having to move towards then (#1).

Another participant noted that their usual learning technique, of writing things down, was not possible in the IVR space (#13).

Implementation

Participants provided various feedback about the system design and its implementation. These included:

- Add a character or person to welcome users (#3)
- Build up from words already known (#3)
- Make it a more structured learning process, less freedom (#4)
- Improve how objects look (#1)
- Offer quick word hints, instead of only the full word (#5)
- Add a way of skipping to the second word on each interaction (#9)
- Think about ways to make it more fun or playful (#17)
- Too long? Better if similar to Duolingo method of five minutes a session (#23)

There was also mixed feedback about the multi-modal aspect of the system, with one participant stating they enjoyed the "coupling of speaking whilst doing" (#8), while another stated that they "don't like learning word and gesture together" (#6).

A few participants offered potential extensions to the experience, such as conversation interactions (#12, #12), using the words being learned in conversations (#17) and creating sentences (#19).

Enjoyment

Ten participants provided feedback regarding their enjoyment in the system, generally regarding the experience being "fun" or "enjoyable". Three mentioned the experience being a "great" method for learning a second language.

4.1.7 Analysis

Environmentally-situated

The broadly positive feedback around the environment suggests that the theorised benefits of environmentally-situated cognition and spatial presence resonate with a large number of participants.

However, mixed feedback around the inclusion of music in the environment raises an interesting question around which aspects of an environment should be recreated in IVR, and which others are not needed or are even potentially harmful. For example, while it is contextually appropriate for a cafe to play music, a participant (#16) noted that it was unusual to study languages with music playing in the background (the included music was a recording of a Japanese radio station playing in a cafe). In the IVR, both being-in-a-cafe and learning are co-occurring, but which one should take priority?

Similarly, there were questions around the suitability of the detail in the environment design, with a few participants reporting that they were distracted by the amount of things to look at. This aspect, however, could be a problem only for users new to games/gaming or to IVR, as these participants reported very low levels of experience with either digital games or IVEs.

A few suggestions from participants regarding improving their experience would make the environment less realist, which again highlights the dichotomy between a 'realistic' experience and an IVR environment designed for learning outcomes. For example, multiple participants mentioned that tea and coffee objects were too similar, and it confused their learning. But tea and coffee objects in the real world, stripped from their smell and taste, can be quite similar. Without the ability to digitise these additional senses, it could be that distinguishing similar objects requites a design choice that presents objects less similar to the physical world but more useful for a learning context.

The same is true of object interaction, where one participant mentioned being able to interact with objects from a distance to "quickly check" what it is called. This type of interaction would collapse the space required to move between objects in the environment, which raises the interesting question - does this space play a useful role in spatial conceptions, or is it just an artefact from the physical-world that could be mediated away in IVR for optimal learning?

Sensorimotor action

Qualitative feedback regarding the sensorimotor activity shows that this aspect was considered enjoyable. Twelve participants took part in the sensorimotor-enabled condition, and half of these, unprompted, stated that this aspect of the system was "fun". While there was no quantitative evidence for motivational differences between the sensorimotor interaction conditions, this feedback certainly suggests that there was at least an enjoyment differences between the conditions; and enjoyment is a motivating factor.

Three participants referred to a sense of connection or reward with the objects they were trying to learn when interacting with them using sensorimotor-enabled interaction - this type of feeling was entirely missing in responses from the non-interactive group.

IVR

Types of locomotion in IVR is a well-studied subject, with some types linked to simulator sickness. The types used in this system were free movement and turning in the physical space, in addition to thumbstick forward/backwards locomotion with snap turning (30 degrees). This combination is typically deployed as one of best setups to avoid locomotion-induced simulator sickness, so it was interesting to note how many people noted issues with it.

While the quantitative results did not show a significant correlation between sensorimotor interaction and cognitive load, one participant mentioned that they used "too much mental energy" to move around and do actions. Therefore it could be that different participants experience cognitive stress differently in IVR.

Participant #13 noted that they typically learn words by writing them down, which was not possible in the IVR. This raises questions about the affordance limitations of IVR implementations, particularly for learning. It is widely understood that people use different learning or memorisation techniques, most of which have been developed in the physical world and are able to be deployed in a variety of situations. However, in IVR, users are constrained to interact only inside that environment, and so cannot, for example, write down words to be learned, unless that option has been programmed into the system. How this limitation of learning options is dealt with inside IVR is an interesting challenge.

It is also possible that participants do not engage in the typical interaction process that is laid out by the system designers. For example, observing one user, instead of enacting the interactive gestures when attempting to memorise the word, or engaging with the virtual object, instead leaned their head backwards and stared at the ceiling while repeating the word aloud. Another user also did not engage with the gesture interactions more than once, instead stepping in and out of interaction zones in order to trigger the audio clip of target word to be learned. Neither of these were engaging with the system in the target way, but were importing the learning mechanisms of visualisation and repetition, which are often used in the physical world.

Implementation

Four participants suggested potential functionality extensions for environment, such as engaging in conversations or using sentences to carry out actions. This task-based learning approach is one that has been linked in literature to IVR and computer-aided language learning generally as being potentially effective, so it is interesting to see it presented by participants without prompt.

4.1.8 Conclusion

This study showed that using sensorimotor-enabled interaction in IVR aided second language memorisation over non-sensorimotor-enabled interaction. It did not find evidence that the use of sensorimotor-enabled interaction had an effect on the perceived cognitive load of participants; their motivation; nor their experience of presence.

The results also evidenced that sensorimotor-enabled interaction and feelings of presence are correlated to learning gain, in ways unrelated to motivation or each other. Through the mediation test, the results also do not support the idea that the contribution of sensorimotor-enabled interaction is measurable or dependent on its impact on motivation or presence. Instead, it seems that sensorimotor interaction should be considered as a unique contributory factor to learning. This then supports the idea that sensorimotor interaction is a "profound affordance" of IVR [141]. Potentially, by enabling deeper levels of sensorimotor-enabled interaction or presence, systems could enhance experiences (in this case, learning outcomes) in IVR.

The distinction between the significant result for verbs and the non-significant result for nouns poses interesting further questions. Could this disparity suggest that the impact of sensorimotor-enabled interaction is limited to the sensorimotor activation of the body, and learning directly related to that activation? If this were the case, learning benefits would not be present for nouns, nor would they be any different when manipulating noun objects as part of the learning interaction process (as opposed to not manipulating objects). This leads to a further investigation: what mechanism inside the sensorimotor interaction condition is causing higher learning gains? Is it because participants' are manipulating their bodies in congruent ways which cause learning - a gesture-learning perspective - or is it because they are able to manipulate the objects in the environment - an action-learning perspective.

The results question how useful motivation works as a metric when recorded inside an already highly motivating experience. For future studies, it may be useful to use participants more experienced with IVR, for whom the experience is less novel (and hopefully, less motivating).

Future work should more comprehensively test the conclusions presented here, ideally using more sensitive measures of presence, and extend the measures of motivation beyond system-level and towards a more subject-specific learning motivation.

The qualitative feedback provided many suggestions for improving the design of learning IVR, but broadly presented high user engagement, enjoyment and acceptance of the system, with a notable limitation of locomotion. In designing further interventions, reducing the amount of locomotion would reduce these issues; and moving users to an abstract learning space would also allow for the exploration of sensorimotor-interaction without the environmental distraction mentioned by some participants. Similarly, a more focused and structured instruction of how to memorise the words may reduce the chance of participants not fully engaging with the sensorimotor aspects of the interaction.

The feedback also highlighted three important conceptual questions regarding IVR learning design:

- 1) How do designers balance the trade-off between creating accurate recreations of environments and optimising them for learning? Music inclusion, environment detail, object manipulation feedback and adjusting the contrast between learning items could all be changes that make the environment less similar to the physical context but provide a stronger learning outcome
- 2) How can designers improve on constraints found in the physical world? Do participants need to be close to objects to interact with them? Can theorised spatial benefits of item locations (such as the benefits afforded by placing objects around an environment, like memory palaces [171]) exist without the burden of traversing that space? How can we improve on learning in reality by using the unique affordances of IVR?
- 3) How do designers address the limitations that being in an IVR puts on our learning affordances? If we are totally immersed in a system, our physical-world methods of learning (such as taking notes) can be impossible unless the system replicates them.

While many of these questions lie outside the scope of this thesis, they provide interesting questions to consider for further research.

4.2 Sensorimotor-enabled interactions in IVR: gestures or actions?

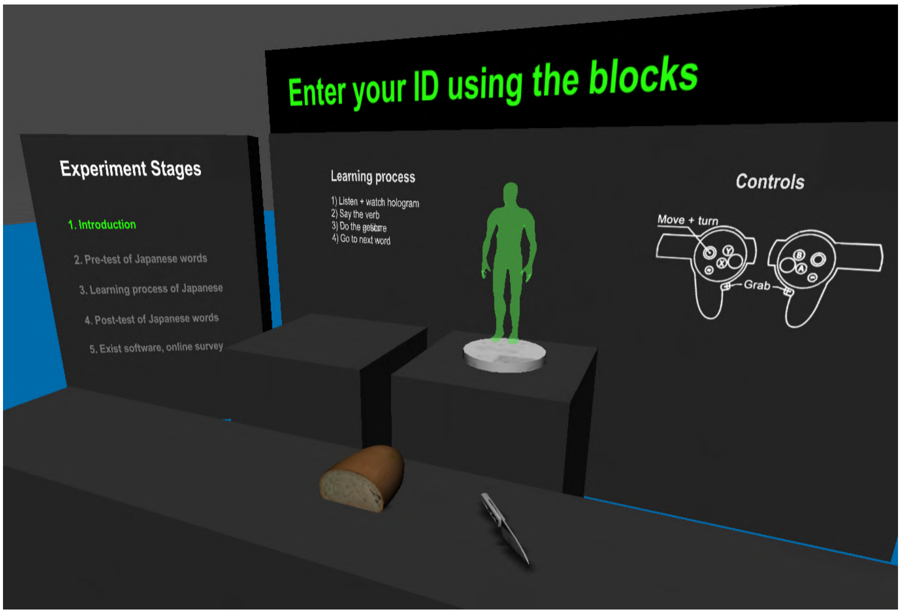


Figure 4.7: Image of the virtual reality environment used in this experiment. Picture shows the more abstract environment design than the first experiment, which includes the stages of the experiment (left), the area in which props appear (currently showing the 'cut' activity, front-centre), and the hologram that acts-out the bodily movement required by the participant (centre)

Hypothesis	Outcome
H1. The action group will demonstrate stronger verb learning gains than the gesture group	Partially rejected
H2. The action group will demonstrate faster response times than the gesture group	Rejected
H3. The action group will report stronger embodiment than the gesture group	Rejected
H4. The action group will report stronger presence than the gesture group	Partially rejected
H5. The action group will report stronger motivation than the gesture group	Rejected

Table 4.6: Summary table of hypotheses and results

4.2.1 Introduction

In the previous experiment, results showed that leveraging sensorimotor-enabled interaction increased learning gain over a non-sensorimotor-enabled interaction. The sensorimotor interactions featured in that system involved participants picking up virtual objects and conducting relevant actions with them (e.g. grabbing a virtual jug and then pouring virtual water from it).

While that experiment treated the process of picking up and using a virtual object as a single HBSI, it is possible to reconsider it as being comprised of two composite HBSI aspects. The first aspect is gesturing, in which participants moved their physical body in ways congruent with their learning. An example of this is participants raising and tilting their arm as if to pour a jug. This activity could take place in IVR or outside of it, as it is dependent on the participant's bodily manipulation.

The second aspect is the manipulation of virtual objects in IVR. Participants moved their body in a way engaged with IVR system, in which they were able to manipulate virtual objects as if they were physical ones. While this requires bodily manipulation, it additionally provides virtual context and feedback in the form of the object being manipulated and responding.

Examining the previous experiment's results from this perspective, it is not clear whether the ability to use physical body activation to interact with a virtual object (referred to from here as actions), provides any benefit over simply activating ones body in a congruent manner without any virtual object interaction (referred to from here as gestures).

Therefore further experimentation is needed to understand if the enhanced learning gain observed in the previous experiment was due to the participants doing gestures, or if they were dependent on actions using the virtual objects as the driver of the enhanced learning gain (see Fig. 4.8 for depictions of gestures and actions in the physical world and in IVR).

If a distinction between gesture and action-based learning outcomes can be found, then there is evidence that what a user is interacting with, while using their body in IVR, is an important factor in how they cognitively process their IVR experience. Therefore, the sensorimotor-enabled impact of IVR interactions would not simply be because our bodies are being activated, but dependent on the virtual stimulus that the bodies are interacting with.

This is important, as there has been little research into the way in which sensorimotor-enabled interactions within IVR are cognitively contextualised. This ambiguity is reflected in literature discussing controllers that facilitate HBSI, in which they are also referred to as gesture controllers or hand gesture inputs [176][180], and sometimes even natural user interfaces [261], despite the

Physical action

Physical gesture

Virtual action

Virtual gesture

Figure 4.8: Images showing actions and gestures in physical and virtual world

use of embodied controllers being quite *un*natural.

Research into sensorimotor activity and cognition in the physical world suggests that body-based actions - interactions with objects - have a different cognitive framework and present evidence of different cognitive outcomes than gestures - body manipulations without objects. Learning with actions, generally, has been shown to make stronger and more specific mnemonic impressions on people experiencing them or enacting them, whether that is for the location of objects [127], or the memorisation of words [334]. They have also been found to be easier for learners to process [153].

If similar learning distinctions between gesture and action conditions are found in IVR, then there is some evidence for an argument that actions in virtual and physical worlds are closely aligned, and that body-based interactions in IVR are not limited to being considered gestures - even if the objects being interacted with are virtual. However, a lack of distinction supports the idea that sensorimotor-enabled interactions in IVR should only be conceptualised through theories around gestures.

In order to determine if this distinction exists, an IVR system for learning Japanese verbs was created, in which half the participants, called the 'action group', learned via actions (being able to use their body to manipulate verb-congruent virtual objects) and half, called the ‘gesture group’, learned via gestures (being able to move their body, but not being able manipulate virtual objects). Learning gain was then compared to see if any difference existed.

4.2.2 Literature

Sensorimotor learning: physical world distinctions between gesture and action

Learning while using sensorimotor gestures or actions has been previously explored in self-performed task (SPT) literature [79][273][63]. Experimental research has shown that both taking actions with objects (SPT-Os) [125] and gesturing without objects aids the memorisation process.

Recent research has provided evidence for a distinction in learning outcome when encoding with either a gesture or an action. Learning with actions, generally, has been shown to make stronger and more specific mnemonic impressions on people experiencing them or enacting them, whether that is for recalling action activities [125], the location of objects [127], or the memorisation of words [333]. They have also been found to be easier for learners to process [153].

Learning with gestures has been shown to promote better representational rather than absolute understanding of objects [226], and an enhanced ability to generalise verbs to wider situations [224][334].

In comparative studies between actions and gestures, Wakefield and Hall [333] found that children learned novel verbs better through action experiences rather than gesture experiences (although they later found similar rates of learning [334]). There have also been higher rates of recognition and recall accuracy for verbs with a greater amount of associated information [292].

There have been numerous explanations for the learning distinctions between action-based and gesture-based learning. The first is that acting-on-objects is cognitively distinct from gesturing-off-objects, and uses different encoding routes, even if the movements are similar [334]. Evidence for this exists in the distinction between physical manipulation theories [208] and gesture-simulated action [126] approaches to embodied cognition.

The second explanation is that the difference in memorisation can be explained by actions increasing the distinctiveness of the memory traces by adding item-specific and relational information [229], compared with gestures. There are two branches of this approach: the first is that actions typically include a corresponding object or target of the action, which brings enhanced multimodal activation (auditory, visual, tactual, olfactory, and gustatory activation) and richer experiential aspects (color, shape, texture, taste, smell, and motor aspects besides the verbal aspects) [9].

The second branch is that the memory traces are dependent not on the object's experiential aspects, but on its action affordances. If we make an action on an object (in the physical or virtual worlds), and the object responds (i.e. is manipulated or moved), we experience the object as manipulable. There is evidence that the perceived manipulability of an object impacts how we memorise it [199]. In this, Madan and Singhal interpreted the overall benefit for highly manipulatable items as being due to automatic activation of motor representations. Perhaps actions-on-objects stimulate the activation of motor representations to a higher degree then gestures-off-objects?

A third explanation is that the enactment effect is not caused be sensorimotor encoding, but by the enhanced "learning episode" the sensorimotor activity creates [167]. By enacting an action like "lifting the pen", the act of lifting and the pen are registered together in a single, specific episode. This is known as the *encoding specificity principle* [325]. This view suggests that actions-on-objects creates deeper episodic integrations than gestures-off-objects. Supporting this view is evidence that semantically sensible learning situations cause stronger memorisation outcomes. For example, Mangels & Heinberg found that semantically sensible action phrases (e.g. "hug the doll") had better memorisation outcomes then stranger ones (e.g. "hug the shovel"), suggesting semantic association played a role in memorisation [205]. Relating this to actions and gestures, perhaps taking actions-on-objects creates a more semantically sensible learning situation than gesturing-off-objects, and hence the noted learning effect.

Sensorimotor IVR: action and gesture distinctions

There are few explorations into the differences between action and gesture encoding in IVR. However, there is some evidence that IVR actions are cognitively processed in a way similar to physical-world actions. Studies into sensorimotor IVR skill development have shown that skill improvements transfer from virtual to physical world domains [111][173][322]. One neuromuscular investigation, into throwing in the real-world and (non-immersive) virtual reality, used electromyography signals of 11 muscles of the upper limbs to examine if conducting a throwing action with an output in a virtual world altered the throwing mechanisms compared with when the outcome was in the physical world. They found very high similarity between the actions in both worlds [278]. However, while the physical mechanisms may be similar, that does not mean the experience or outcome are the same. For example, another study found that throwing precision and accuracy in IVR are lower than in the physical world, and that the throw action in IVR requires more user effort, and, slightly contradicting the

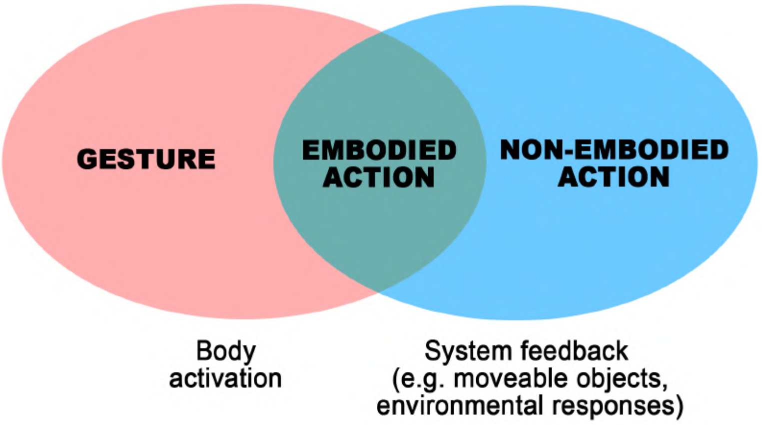


Figure 4.9: Diagram showing proposed distinctions between gestures and action in IVR. Actions (both embodied and non-) feature two system-created feedback points: interactional (user is able to move objects) and environmental (the world responds to the user's movement of objects). The overlap of the system feedback and body activation is embodied action, which in this paper is compared with gesture

above study, produces a different kinematic throwing pattern [354].

Investigations into the IVR body transfer illusion [298] provide further evidence that virtual actions could be cognitively perceived in a similar fashion to physical world actions. According to research in this area, users perceive the actions of other agents on their virtual bodies in a similar way to physical actions on their physical bodies. However, it is important to note that this research is about responding to an action from another, rather than self-performing an action.

While much more research is needed, particularly for the cognitive outcomes based on IVR actions (rather than simply the actions themselves), the existing findings suggest that actions can take place in IVR that are similar to those in physical world. And if, as embodied cognition theory suggests, the boundary between a cognitive agent and his or her environment can be considered malleable [4], it follows that IVR actions that have similar outcomes to those of physical world actions may encourage our brains to think we are taking actions in IVR, and not just outputting gestures.

4.2.3 Experiment

This between-subject experiment investigates if there is a distinction between encoding with actions or gestures on verb memorisation in IVR. For the action condition, participants were able to use objects to complete actions in order to learn a congruent verb. For example, a participant had to pick up a cup, bring it to their mouth and tilt it while learning the Japanese word for 'drink'. In the gesture condition, participants had to make a gesture relevant to the verb, but were unable to interact with the object (i.e. could not touch or move the cup). In both conditions, an animated 3D model also demonstrated the action/gesture for the verb.

The learning gain of participants in each condition was compared to understand the role that interactions play in cognitive and memorisation of verbs. Response times for correct answers were also monitored, as faster responses times can be representative of depth of learning. Participant embodiment, presence and motivation scores were also monitored to allow for the investigation of potential correlations between these metrics, interaction conditions, and learning gain.

Hypotheses

The below hypotheses are based on literature that presents actions as more powerful verb encoders than gesture. Additional hypotheses suggest that distinctions between actions and gestures might reflect in affective factors related to IVR, such as embodiment, presence and motivation results:

- H1. The action group will demonstrate stronger verb learning gains than the gesture group
- H2. The action group will demonstrate faster response times than the gesture group
- H3. The action group will report stronger embodiment than the gesture group
- H4. The action group will report stronger presence than the gesture group
- H5. The action group will report stronger motivation than the gesture group

Procedure

Participants were asked to download and run an executable file on their personal IVR-enabled computer systems. An on-boarding process, pre-test, learning process and post-test all took place within the downloaded software.

After opening the software, participants were assigned to one of two interaction groups, which they were not aware of: action or gesture. They differ as follows:

- Action: Input is made by grabbing a virtual object, and doing the correct action with it. A completed action is given some kind of feedback (e.g. drinking sounds for the action of drinking from a virtual cup)
- Gesture: Input is made by doing the correct gesture in the air, away from the object. The virtual object is presented but not manipulable.

Participants were given an on-boarding process that explained the control methods (both aurally and written), and required the participants to move to a target location and touch an object to continue the experience. A voice-over then explained the experiment goals and four-stage process: tutorial, pre-test, learning, post-test.

The tutorial demonstrated how the learning process worked. This involved demonstrating and explaining the process with a word and object that was not included in the actual learning and testing process. It explained that they should listen to the audio explaining the word to learn (i.e. "Eat the doughnut. Eat is taberu. Taberu. Taberu"); watch the 3D model act-out the eating process, then repeat the word aloud ('Taberu') and complete the gesture without trying to touch the object, or pick-up the object and do the action. The final part of the instruction depended on the participant's assigned condition.

Participants were then pre-tested for their knowledge of 15 Japanese verbs (the rationale for changing this from 10 verbs and 10 nouns as in the previous experiment can be found in Chapter 4.2.3). The pre-test involved listening to a Japanese word and choosing its English meaning from a list of 15 verbs, or skipping the question. Answers selected by participants were not removed from the list of possible answers, nor was feedback given whether their selection was correct or not, so all questions had 15 possible options, and participants could repeat answers. Questions were presented sequentially, and participants were not allowed to amend previous answers.

During the learning process, participants were asked to memorise 15 verbs (see Table 4.7 for list). Participants were exposed to each verb in sequence (1 - 15), for five sequences. Each verb related to a different object presented in front of the participant on a podium in the IVR. Participants were told an action/gesture, the English verb, and the Japanese language verb. For example, a phrase used for learning *drink* was "drink from the cup. Drink is nomu. Nomu. Nomu". This phrase was used for both the action and gesture condition, with participants either doing the drinking action by manipulating the 3D model of a cup, or doing a drinking gesture without manipulating the 3d model. A 3D animation of a humanoid model gesturing (i.e. moving but not manipulating an object) was also displayed for both conditions.

For each verb, the participant had to either gesture or action once (depending on their group) and say the verb aloud once. Participants were instructed to say the verb aloud in order to control for the Production Effect, in which speaking a word while encoding it causes stronger memorisation than not speaking it [197]. Both groups of participants used the same avatar - a set of white hands with no arms or body.

After the learning process, participants took a post-test, which was the same procedure as the pre-test. Learning gain was calculated as the final test result minus the pre-test results, and normalised into a number between 0 and 1. A further web-browser-based test was taken one week after the initial study to determine learning retention.

As this was not a controlled, lab-based experiment, due to the impact of Covid-19 on research procedures, the data collection was done inside a VR environment. To verify the gestures and actions were completed correctly, telemetry of participant movements was recorded and examined.

Participants

Fifty-six (56) participants took part of in the study. Of these, 53 were compensated and three were uncompensated. Uncompensated participants volunteered to take part after compensation offers had closed and this change had been advertised.

Forty-eight (48) participants' data was usable in the analysis. Two participants were excluded for having high levels of pre-existing Japanese knowledge (they already knew six and eight of the 15 target words). One participant was excluded for presenting unusual movement data. A follow-up conversation revealed they were using a spoofed virtual reality system (i.e. they used a monitor, mouse, keyboard and emulator to access and play VR content). Five participants were excluded due to incomplete data being returned from the remote software, and not manually forwarding the data when requested.

All valid participants who reported their recruitment referrer (38 participants) came from an unpaid advertisement posted on the Reddit */r/oculus* community and used their own IVR hardware in their own setting. This suggests the participants were experienced in using IVR hardware.

The average age of valid participants was 27 ($SD = 6.75$). Participant gender skewed heavily male (38) over female (8) or other/did not say (2). Valid participants had a low knowledge of the target learning words during the pre-test, with the average participant knowing less than one word ($M = .15$ (out of 15); $SD = .46$).

The majority of valid participants were fluent in more than one language (17 self-reported as fluent in one language, 26 self-reported as fluent in two

Verb	**Action**	**System feedback**
Wear	Pick-up a hat and place on head	Hat sticks to head, appears at top of vision
Wash	Pick-up a plate and place in a sink	Plate submerges in water, washing sound plays
Drink	Pick-up a cup, bring to mouth and tilt	Drinking sound plays
Smoke	Pick-up a cigarette and bring to mouth	Inhaling and exhaling sound places
Climb	Place hands on vertical climbing rope	Player is raised into the air as if climbing
Open	Pick-up a a box lid from a closed box	Box lid makes a noise on grab and put-down
Grab	Pick-up a bank note from a table	Money makes a noise on grab and put-down
Take (a photo)	Pick-up a camera and point it at a dog	Camera makes a shutter noise when facing dog
Press	Push down on an industrial button	Button compresses when pushed
Pull	Grab rope, pull away from fitting	Rope extends as if pulled out from fitting
Turn on	Push hand into lightswitch	Lightswitch gets depresses, makes clicking noise
Raise	Pick-up an umbrella and hold above head	Raindrops are blocked by umbrella
Brush	Pick-up toothbrush and bring to mouth	Brushing sound is played
Set/place	Pick-up a cup and place on a tray	Cup makes a noise on connection with tray
Cut	Pick-up knife and moved into bread	Slice of bread is cut from loaf, makes noise

Table 4.7: List of target words, the participant's actions for encoding (for the action condition) and the feedback given by the system when an action was successfully completed. For gesture participants, the participant were instructed to move their hands in the same way as if they were carrying out the action. All actions and gestures were presented via an animated hologram person in front of the participant

languages, 4 in three languages, and one in four languages). No significant correlation between languages known and learning outcome was found ($r = 0.24$, $p = 0.10$).

Interaction condition was randomly assigned inside the software. As such, 27 participants were assigned to the *action* condition and 21 to the *gesture* condition.

Corpus

Participants were tested on their knowledge of 15 concrete verbs. Concrete verbs are verbs that represent very literal actions, such as run, jump, throw. These were chosen as they are highly embodied and were used in previous gesture and action word memorisation comparisons [334]. The target words were chosen to be familiar actions that allowed for mostly distinct body manipulations for each word.

Unlike the previous experiment (Chapter 4.1), this experiment did not feature noun learning. This is because the experiment is investigating whether any distinction between actions and gestures can be found, and not whether it can be found in all cases. As the previous experiment was able to find a significant result for verbs, and because we are investigating sensorimotor activities, it was felt that an effect would be more likely to be detected if using verbs as the learning subject.

Japanese gairaigo (import words) were specifically avoided to reduce the chance of participants' inferring a meaning from their similarity to English. Additionally, efforts were made to reduce the use of phonetically similar and particularly long Japanese words, as there were concerns that beginner-level learners would find these words difficult to tell apart. A list of the words used, the participant's action (for the action condition) and the feedback given by the system can be found in Table 4.7

Environment

A 3D environment created in Unity (2018.4.9) was used, based upon the Unity WebXR Exporter project by MozillaReality for handling IVR interaction (note: although Unity WebXR Exporter is designed to convert Unity XR content to the web, it was not used that way in this case, with the experiment running as an exectuable file on the participants' computer). Based on the learnings from the previous experiment (Chapter 4.1), the decision was made to not use a realistic aesthetic and layout, but an abstract space that minimised potential distractions (both visual and aural), reduced the need for player locomotion around the space, ensured objects were presented close to the player, and maximised the use of supporting text and prompts to guide the experiment process.

The objects presented to the participant as part of the learning process were presented in a realist style, although with reasonable limitations to their texture and polygon details to avoid impacting performance (under one million triangles per object).

The environment was explorable via a head-mounted display and embodied controllers. Participants used their own IVR hardware to take part, and where reported, these were overwhelming Oculus/Meta Quest headsets and controllers (95%). This is almost certainly an artefact of the popularity of the headset and of the recruitment process. Automatic reporting of hardware was not included in the telemetry data.

Participants could navigate the virtual environment by moving around the real space and/or by using the thumbsticks on the controllers. The left thumbstick was used for forwards/backwards locomotion and strafing; the right thumbstick for 'snap' turning at 30 degrees at a time, in either direction. The movement speed was dependent on how far forward or backward the left thumbstick was pressed, but subjectively may be described as slow. This speed was chosen in order to reduce the risk of locomotion-induced simulator sickness. Participants could only pick-up objects by moving their hand so that the hand graphic was touching or embedded in the virtual object, and pressing the 'Grip' button on the Oculus Touch or Vive Wand controllers. The objects did not snap into a fixed position in the hand.

Participants' movement data was logged at intervals of 0.1s (10Hz). Movement data included head and hand positions and orientations. This frequency was selected after trialling different resolutions, examining their total log size, and determining what was a reasonable and reliable log size to send to the experiment servers. The 10Hz frequency created a log file (dependent on session time) of around 2MB, which was well-tolerated by the server receiving the data. Action/gesture completion rates, and time per action/gesture, were also recorded in order to identify potentially problematic interactions or sessions.

Evaluation

Participants' knowledge of the verbs was measured in three tests: one administered before their exposure to the environment (pre-test); one immediately after (post-test), and one seven days later (week-test). Participants performed the same test each time, listening to a Japanese word and choosing the English meaning from a list of 15. All three tests were conducted remotely outside of laboratory conditions; the first two were conducted inside IVR and the third was via a web browser. The time taken for each question was recorded to help us evaluate the testing sessions. A visual examination of this data did not highlight any individual user taking a consistently long or short time to answer each question, suggesting that participants avoided looking-up answers; being consistently distracted (in a way that could be measured by time) during the evaluation; or rapidly entering answers in order to receive payment. Participants were not given feedback when submitting answers.

Learning gain was calculated as a normalised score between 0 and 1, measured as post-test divided by the number of eligible words for their session. Five participants had existing knowledge of either one or two of the target verbs. In these cases, the words were removed from their post-score, pre-score and eligible word totals:

$$(PostScore - PreScore)/(EligibleWords - PreScore)$$

Whether participants listened to the audio clip before submitting an answer was tracked, which was the case for every entry except one, who missed one question. It is believed that this was the result of an accidental double-input on the previous question, and so this word was removed from the participant's score and eligible word total when calculating the normalised result.

After using the system, participants were asked to complete a survey in-browser. This consisted of the Gonzalez-Franco & Peck immersive VR embodiment questionnaire [103], the Igroup presence questionnaire (IPQ) [282] and the Intrinsic Motivation Inventory intrinsic motivation questionnaire (IMI) [268].

The Gonzalez-Franco & Peck embodiment questionnaire was chosen as it is the only attempt at a standardised embodiment questionnaire for IVR research that could be found. Its division of embodiment into six sub-scales for differing experimental interests allowed for the isolation of a form of embodiment particularly relevant to this research: agency and body ownership. Participants were presented with statements regarding their sense of embodiment and asked to mark if they "strongly disagree", "disagree", "somewhat disagree", "neither agree nor disagree", "somewhat agree", "agree" or "strongly agree" on a seven-point scale (from -3 to 3).

The IPQ was also chosen due to its ability to measure sub-types of presence. All four types were included - general presence (the "general sense of being there"), spatial presence ("the sense of being physically present in the IVR"), involvement (measuring "the attention devoted to the IVR and the involvement experienced"), experienced realism (measuring "the subjective experience of realism in the IVR") [282]. Each of these types could have implications for the cognitive perception of actions and gestures in the IVR. Participants were asked to provide ratings on statements on a seven point scale (from -3 to 3). The survey contains one question on general presence (and thus the minimum possible general presence score was -3, and the maximum possible was 3), five on spatial presence (minimum possible score -15, maximum possible score 15), four on involvement (minimum -12, maximum 12) and four on realism (minimum -12, maximum 12).

IPQ is also a well-validated method [202][330]. Asking participants for their evaluation of presence experienced is considered the most direct way to assess presence [135].

The IMI is a well-established tool for measuring sub-scales of motivation [212]. Because this is a learning experience, the interest/enjoyment and value/usefulness sub-scales were used. Participants were asked to rate how true they felt statements regarding their experience were on a seven point scale, from “Not true at all (1)” to “Very true (7)”. This survey replaced the MEEGA+ survey [243] used in the previous experiment (Chapter 4.1). This was because the Intrinsic Motivation Inventory has been more widely validated, and captures feedback concerning the learning session experience, rather than attempting to focus on the learning system itself.

The questions given to participants for the immersive VR embodiment questionnaire, the IPQ and the Intrinsic Motivation IMI questionnaire [268] can be found in Appendix B.

Participants’ telemetry data was examined in order to determine if participants had issues taking part in the study. The process for this first identified the time to complete each action/gesture, with any process taking above 20 seconds marked for detailed exploration. For any sessions identified, the participant’s movement data was played back and examined inside the environment, to attempt to understand if something had gone wrong in the process. Only a handful of action/gestures took over 20 seconds (n ¡ 10, out of 3600 potential action/gesture interactions), and of those, it appeared that head and hand positions stopped for extended periods during that particularly action/gesture, suggesting either a technical issue or a focus on something outside of the IVR hardware.

Cognitive load was not measured in this experiment. This was predominantly because no impact of cognitive load was found in the previous experiment, and it was believed that the previous experiment was more likely to present a cognitive load distinction than this one, due to a more notable change in sensorimotor activity between the interaction conditions. If, in the previous experiment, there was a distinction in cognitive load, but that distinction was not detected by the self-reported measurement used, it was believed that using the same measurement approach in this experiment would be insufficiently sensitive to detect a distinction. A preferable solution would be to measure cognitive load using physiological indicators, but as this study was conducted with remote participants, this was infeasible. There was also desire to limit the number of questions given to the remote participants.

Analysis

The first hypothesis ("does the actions group demonstrate stronger verb learning gains than the gesture group?") was tested by coding correct responses as 1 and incorrect responses as 0. Where a participant had answered correctly in the pre-test, their future responses for that word were removed.

A Mixed Models was used to account for both the fixed (interaction type) and potential random (users, words) effects, as recommended by Macedonia et al. [194]. This approach is more robust than in the previous experiment (Chapter 4.1), as the model takes into account random factors outside of the primary area of concern (interaction type), such as variance in participants and words. As the dependent variable was binomial, it was a Generalised Linear Mixed Model.

To test the second hypothesis (the actions group will demonstrate faster response times than the gesture group), a Linear Mixed Model was used due to the continuous dependent variable of response time. Only correct answers were included in the dataset, and outliers were removed. Outliers were highlighted by checking for 1.5 * interquartile range above the third quartile, or below the first quartile. These outliers were split fairly evenly between conditions, and some participant's mean answer times were skewed by a few longer entries, so it was felt that removing these would increase accuracy of the results by more accurately reflecting participant thinking time, instead of an artefact of the testing process, with a low risk of biasing the results for one conditions.

For the third hypothesis (the actions group will report stronger embodiment than the gesture group), linear regressions between each of the two surveyed embodiment scores (ownership and agency) and the interaction condition (action or gesture) were used.

For the fourth hypothesis (the actions group will report stronger presence than the gesture group), linear regressions between each of the four presence scores (general, spatial, involvement and realism) and the interaction condition (action or gesture) were used.

For the fifth hypothesis (the actions group will report stronger intrinsic motivation than the gesture group), linear regressions between motivations scores (interest, value/usefulness) and the interaction condition (action or gesture) were used.

4.2.4 Results

The comparison of pre-test results of included participants found no significant difference ($t = 1.31$; $p = .20$) between the pre-existing knowledge of the action group ($M = 0.01$) and gesture group ($M = 0$), with only five participants knowing any Japanese (means presented as normalised from 0 to 1).

H1. The actions group will demonstrate stronger verb learning gains than the gesture group

Results	**N**	**Normalised Mean Score**	**Mean RT**
Action: Post-test	27	0.66 ±0.27	9.13 ±4.84
Gesture: Post-test	21	0.47 ±0.25	8.90 ±3.92
Action: Week-test	21	0.39 ±0.25	5.76 ±4.53
Gesture: Week-test	14	0.27 ±0.19	5.35 ±4.85

Table 4.8: Table of learning gain results (normalised mean) and response time from tests immediately after the session (post-test) and one week later (week-test), with standard deviation

Parameter	**Beta**	**Lower-95**	**Upper-95**	**Std. Error**
Post-test				
Intercept	-0.13	-0.90	0.64	0.38
Interaction (Action)	1.12	0.24	2.04	0.44
Week-test				
Intercept	-1.30	-2.04	-0.63	3.46
Interaction (Action)	0.79	-0.03	1.64	0.41

Table 4.9: Table of Generalised Linear Mixed Model results for learning gain. Note: co-efficients are logit

Parameter	**Beta**	**Lower-95**	**Upper-95**	**Std. Error**
Post-test				
Intercept	9.32	7.92	10.76	0.71
Interaction (Action)	0.07	-1.35	1.47	0.70
Week-test				
Intercept	5.82	4.19	7.48	0.81
Interaction (Action)	0.11	-1.28	1.54	0.70

Table 4.10: Table of Linear Mixed Model results for response time

The descriptive results for both post-test and one-week learning gain (and response times) are presented in Table 4.8, and the GLMM results are presented in Table 4.9.

For the post-test, the GLMM (n = 720; 48 participants) showed learning gain varied across both participants (σ^2 = 1.88) and words (σ^2 = 0.58) After controlling for these random factors, the model presented a statistically significant relationship between interaction type and learning gain (p = .012). Words encoded in the action group were better remembered than those in the gesture group (β = 1.12, 95% CI [0.24,2.04]). In the model, given a participant and word with average intercepts, if they were assigned to the action condition, they would be 26% more likely to correctly remember a word than in the gesture condition (73% vs 47%).

For the one-week follow-up test, the model (n = 525; 36 participants) also showed learning gain varied across both participants (σ^2 = 1.00) and words (σ^2 = 0.22), but likely to a lesser extent. It did not present a significant distinction in learning gain between words encoded in the action group (β = 0.22, 95% CI [-0.03,1.64]) and those in the gesture group. Although not significant, in the model, given a participant and word with average intercepts, the probability of getting a correct response increases from 21% to 37% in the action group.

Therefore H1 is accepted for the immediate post-test results, but not for the week-test.

H2. The actions group will demonstrate faster response times than the gesture group

The descriptive response time results for both post-test and one-week test are presented in Table 4.8, and the LMM results for both post-test and one-week later test are presented in Table 4.10.

For the post-test, the LMM (n = 402; 48 participants) showed response time varied across both participants ($\sigma^2 = 3.66$) and words ($\sigma^2 = 3.07$). After controlling for these random factors, there was no significant distinction between between the action group ($\beta = 0.07$, 95% CI [-1.35,1.47]) and the gesture group.

For the week-test, the model (n = 173; 33 participants) showed response time varied across words ($\sigma^2 = 4.47$) and to a lesser extent users ($\sigma^2 = 0.06$). After controlling for these random factors, there was no significant distinction between between the action group ($\beta = 0.1$, 95% CL [-1.28,1.54]) and the gesture group.

Therefore H2 is not accepted for either immediate post-test response times or the week-test.

Words by InteractionType

After finding some evidence of the random effect of words on learning outcome, a LMM was used to explore whether words had an interaction effect with interaction type; to understand if some words were better or less suited to embodied encoding. However, a likelihood ratio test indicated that adding random intercepts for each interaction condition of each word (word*interactionType) did not improve the model over adding random intercepts for each word only. Therefore a significant interaction between word and interactionType was not found.

H3. The actions group will report stronger embodiment than the gesture group

The results did not show a significant correlation between the interaction type (action or gesture) and the self-reported feeling of embodied agency ($r = 0.64$; $p = .67$).

However, four potential outliers (based on inter-quartile range) were observed, and when these were removed, the results showed a significant correlation ($r = 0.33$; $p = .03$), which would mean interaction type explains 10.8% of the variability of the embodied agency score. A graph depicting the linear correlations between embodied agency and interaction type, both with and without outliers, is presented in Fig. 4.10.

It was difficult to determine whether these were true outliers or not. These four participants presented embodiment ratings that appear distinct from their peers, however it is not impossible for them to have felt incredibly embodied (or non-embodied) by the interactions, or to have interpreted the question notably differently from others. One of the outlier participants entered universally the lowest scores for all embodied agency questions, but provided more varied results for other questions.

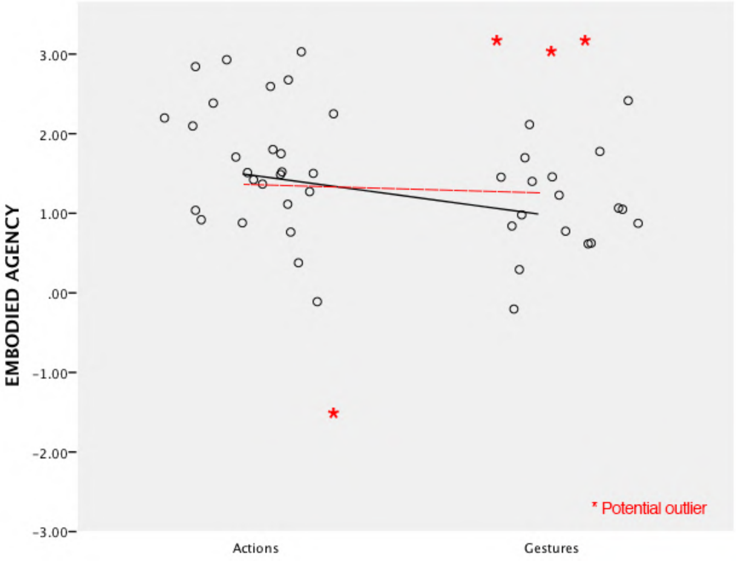


Figure 4.10: Jittered plot showing relationship between self-reported embodied agency and interaction type. Black line shows significant relationship when outliers removed, red line shows relationship without significance when outliers are included.

Without strong evidence to remove these outliers, however, they have been included them in the dataset, and so it is not possible to report a significant relationship.

The relationship with body ownership ($r = 0.05$, $p = .71$) was not significant.

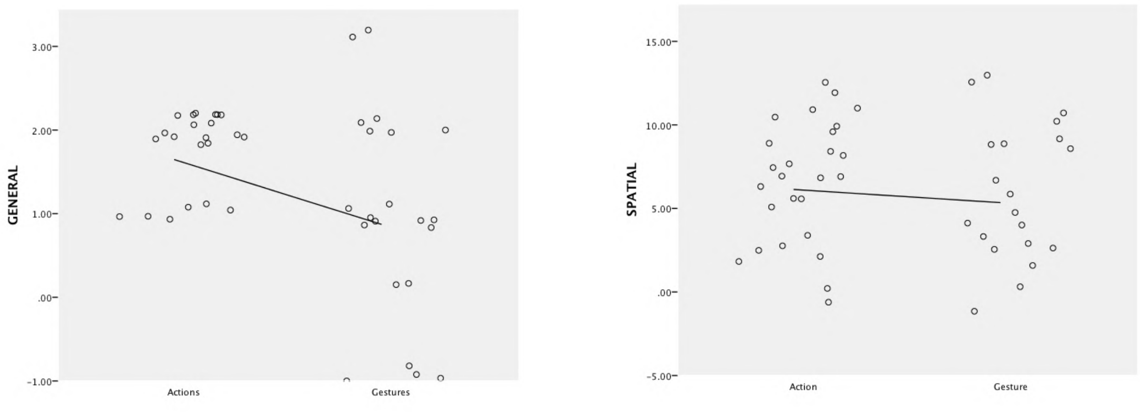
Therefore H3 is not accepted.

H4 The actions group will report stronger presence than the gesture group

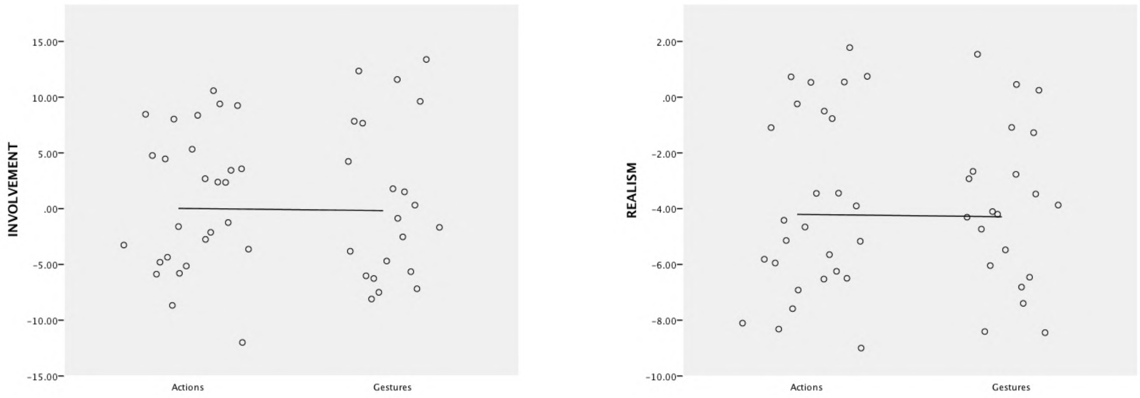
Presence Type	Action Mean	Gesture Mean	R	R2	P
General	1.7 ±1.1	1.0 ±1.2	.32	.103	.026*
Spatial	6.0 ±4.3	5.0 ±5.2	.01	.012	.045*
Involvement	0.6 ±5.9	0.4 ±6.8	.02		.091
Realism	-4.0 ±3.2	-4.0 ±2.9	.01		.092

Table 4.11: Summary of presence scores and the size and significance of their relationship with interaction type.

The results showed a significant correlation between the interaction type and



General Spatial



Involvement Realism

Figure 4.11: Graphs depicting linear correlations between interaction type and presence scores, arranged by presence types (general, spatial, involvement and realism). There were significant relationships for general and spatial presence, with a small influence on spatial presence score. The total score for each participant is presented on each presence graph. As different presence types had different numbers of questions, the minimum and maximum of each type varies

self-reported general presence ($r = 0.32$, $p = .026$), which means interaction type explains 10.3% of the variability of the general presence score.

The results also showed a significant correlation between the interaction type and spatial presence ($r = 0.01$, $p = .045$), which means interaction type explains 1.2% of the variability of the spatial presence score.

There were no significant correlations between interaction type and the involvement presence score ($r = 0.016$, $p = .091$) or realism presence score ($r = 0.014$, $p = .092$).

A summary of the presence scores is presented in Table 4.11, and linear correlations for each of these presence measures and the two interaction types is presented in Fig. 4.11.

Therefore H4 is accepted for general presence and spatial presence, but not for the involvement or realism presence variations.

H5 The actions group will report stronger motivation than the gesture group

There were no significant correlations between interaction type and interest motivation (r = 0.049, p = .741) or value/usefulness motivation (r = 0.013, p = .931).

Therefore H5 was not accepted.

4.2.5 Discussion

Evidence sensorimotor-enabled interactions are actions, not gestures

The results show that verb learners who take actions on objects in IVR achieved significant and large memorisation gains over learners who make gestures without manipulating objects. This was reflected in immediate learning gain scores, but not by response times. These results have obvious implications for designing optimal IVR-based action-verb learning applications, which should activate users sensorimotor-systems in a form that includes objects for the interaction and feedback from the system for the object's congruent use.

The explanation for why these results were observed, and what that means for sensorimotor-embodied control in IVR, is more nuanced. It is possible that the learning gain differences between action and gesture conditions in the IVR can be explained by the same cognitive phenomena that has previously been evidenced in physical world comparative studies between action and gesture. If Wakefield's explanation of different encoding pathways between actions and gestures in the physical world [334] is true, then this could be the explanation for the results presented here.

Extending this further, it would mean that sensorimotor interactions using controllers in IVR provide a cognitive experience similar to that of sensorimotor interactions in the physical world: interacting with objects in the physical or virtual worlds are actions; while activating the body to make movements that do not interact with objects, in the physical or virtual worlds, are gestures. Therefore when considering the cognitive aspects of interactive sensorimotor actions in IVR, discussion should stem from the perspective of action-based embodiment theory, rather than gesture-based theory. In short: the cognitive experience from sensorimotor interactions in IVR are based upon what the users are experiencing in IVR, and are not limited to only being influenced by the bodily activation.

However, there is also another potential explanation for the results found here, which stems from the explanation of the enactment effect as enhancing memory traces from multi-modal encoding [229], or from the benefits of enhanced "learning episodes" [167]). The results could be demonstrating the added learning efficacy that stems from the contextually-deployed system feedback and richer situational encoding offered by the action condition. The feedback is two-fold: first from the virtual objects being able to be moved, and second from the system responding to user's manipulations of objects with sound effects or system events. If this was the explanation, it would be inappropriate to extrapolate whether sensorimotor interactions through embodied controller are cognitively considered actions or gestures; and even whether that distinction was meaningful, given that the enhanced learning is (in these perspectives) a result of thorough memory traces or learning episodes, rather than a property unique to sensorimotor encoding pathways.

The secondary explanation lead us into an interesting place regarding the simulation of actions in IVR, in that the amount of "memory trace" could be adjusted for exploration in a way not possible in the physical world. For example, in the physical world, it is unlikely that you can separate actions from their contextual environmental feedback - pouring a jug of water will always cause water to fall (unless, perhaps, you are in space). However, in the IVR space, we are able to have different forms of action-feedback that do not correspond to the real world. Whether water falls, how it falls, does not fall, floats upwards or even exists as at all are the choice of the systems' designers.

An interesting additional exploration to provide further clarity on the cause of encoding differences in the IVR would be to amend environmental feedback to gestures (the gesture for pouring water would not move the object, but would play a pouring sound), or stripping environmental feedback from actions (you could move a jug to pour water, but no water or sound runs out from the jug), and contrast these results to our virtual recreations of typical physical world gestures and action processes. We may find that both interactional and environmental feedback are needed for us to contextualise sensorimotor embodied controller actions in IVR in a way similar to physical actions, or that the benefits can be added to gestures through environmental feedback.

Missing retention, similar response times

There was no evidence of a significant difference in learning retention after one week between the two encoding conditions, although the action condition showed a higher, non-significant, mean learning gain. This is similar to results in the previous experiment (Chapter 4.1). There are three potential explanations for the difference in significance between the immediate and one-week later tests: (1) the drop in participants (from 48 to 35), as many did not complete the one-week later test, reduced the sensitivity of the test; (2) the difference in learning gain between the conditions is reduced after one-week due to a general drop in learning gain, thus reducing the sensitivity of the test; (3) learning gain differences between action and gesture only occur immediately, and longer-term learning is similar between conditions. The most likely explanations for the results presented here are (1) and (2), as a reduction in the learning difference between experimental groups over time is a pattern familiar in language learning investigations [267][194]. In this experiment, the learning gain for action dropped 41% between the post-test and the week-test; and 42.6% for gesture, making it harder to detect a significant difference.

There was no significant distinction in response times between the action and gesture conditions. As faster response times are typically associated with stronger encoding [194][90], it is expected to see faster response times for the action condition to match the learning gain scores. However, given that the difference between the learning gain of the two conditions was so large, and that response time is a less direct measure of learning outcome, then the lack of a significant result here does not impact the evidenced learning gain distinction between the two conditions.

Presence and embodiment

The results show that participants in the action condition had significantly higher feelings of general presence and spatial presence (albeit the latter with a small correlation), but not involvement or realism. These results suggest that being able to interact with objects in a virtual space enhances the sense of being present in the IVR. As the experiment was targeting learning in an abstract space, it could be that no distinctions between involvement or realism were found due to the already involved nature of any learning process, or the unrealistic environmental setting.

It was a little surprising to not find any significant relationships between the interaction types and the chosen embodiment measures of ownership or agency. It seems reasonable to assume that of the subjective measures explored here, these would be the most likely to be affected by the different interaction types; as well as higher levels of self-reported embodiment for the interactive object-manipulation. This result presents questions over the relevancy or efficacy for this embodiment survey [103] for this type of exploration of sensorimotor embodiment. The agency-related questions in the survey ask about visuo-motor synchronous stimulation (e.g. "It felt like I could control the virtual hand as if it was my own"), which according to the results, appear to be experienced consistently whether you are interacting with virtual objects or not. Perhaps embodiment in a virtual body, and how that body can interact with the virtual space, is an additional factor that needs a separate survey categorisation.

4.2.6 Limitations

The study participant demographics are a notable limitation of this study, as participants were used who were both familiar with IVR (enough to own their own headset) and had a large enough interest in the technology that they were members of an online community for it, from which they were recruited. A major implication for this is that the audience might be self-selecting: those who IVR resonate with are potentially more likely to have invested in the hardware than the general populous, and so may be more keenly affected by its affordances. The sample was also heavily skewed towards men, who have been shown as less likely to suffer from simulator sickness with the current incarnation of IVR technology [303].

There are also limitations to the generalisability of this research to other uses of IVR. For example, it is not clear if the evidence presented here for the similarity of benefits between action-based learning in IVR and in the physical world, would work for other academic subjects (e.g. mathematics) or other areas (e.g. empathy-training, rather than cognitive learning).

Additionally, this study uses highly sensorimotor-embodied words: concrete verbs. Further study of words more peripherally linked to actions, such as nouns, adjectives, and abstract or stative verbs are needed - although this is also true outside of IVR investigations (existing research suggests that while "the sensorimotor neural network is engaged in both concrete and abstract language contents ... concrete multi-word processing relies more on the sensorimotor system, and abstract multi-word processing relies more on the linguistic system" [272]).

The method for assigning participants to conditions could also have been improved. For the experiment, a random assignment was given on opening the software, giving each participant an even chance of being in either condition. This approach was taken to reduce the dependency on a centralised server. However, over the participant population, this ended up being quite uneven (27 action and 21 gesture). Using server-assigned conditions could have ensured an even participant distribution between conditions.

4.2.7 Conclusion

These findings show that users can have distinct learning outcomes from HSI in IVR depending on whether those interactions were presented as actions (i.e. were able to interact with objects) or gestures (i.e. were not). Learners who encoded information while doing actions had significantly better learning outcomes than those who encoded with gestures.

This suggests that sensorimotor-enabled interaction allows users to have cognitively distinct experiences in IVR depending on what they are doing in the virtual environment (i.e. if they are interacting with an object or not); and that sensorimotor-enabled inputs are not just ‘gestures’, but depend on how the interaction is designed in IVR.

As the action group provided better learning benefits, in the same way as in physical world studies, these results also suggest that IVR users cognitively respond to actions in IVR in a similar way to physical world actions, and to gestures in IVR in a similar way to real-world gestures. If one subscribes to the view that humans memorise information differently depending on whether it was encoded using an actions or gesture, these results could mean that participants had cognitive experiences in IVR that were similar to physical world actions and gesture experiences. This suggests that our cognitive perceptions of interactions in IVR are not restricted by the controllers or abstracted physical world bodily movements, but by what we are experiencing inside IVR.

While this has only been evidenced for learning (in this study), if this were the case generally, it would mean that we should consider sensorimotor actions taken in IVR in a similar way to how we contextualise actions in the physical world, and therefore these interactions should not be limited to being discussed from a gesture-interaction framework. This could have implications for the emerging use of IVR in PTSD or exposure therapy.

However, the observed learning difference could also be explained by theories around memory traces and multi-modal encoding depth. From this perspective, the actions in IVR provided additional interactive feedback (objects could move and make sounds), which the gestures did not. If this explanation is, in fact, the correct one, then it is more difficult to outline a strong case for how we cognitively contextualise our HBSI in IVR. Further research into whether it is the sensorimotor-enabled action, or the additional memory traces of object interaction, or both, that is causing the enhanced learning outcomes evidenced here.

4.3 Sensorimotor-enabled interactions and virtual object feedback

Figure 4.12: Image of the virtual reality environment used in this experiment. Picture shows the abstracted environment design similar to the previous experience, with the interaction "hammer the nail" for learning the verb 'hammer'

4.3.1 Introduction

The previous experiment demonstrated that previously evidenced language learning differences between non-IVR gesture-based and non-IVR action-based encoding persist into IVR. This result also provides evidence for the argument that gestures and actions are processed as cognitively distinct sensorimotor activities in IVR, as they are theorised to in the physical world. However, it is not entirely clear if the distinction in learning gain is dependent on learners processing gestures and actions in distinct ways (as theorised, regarding physical world actions, in [334]), or if the object manipulation enabled by actions simply provides more modes of feedback, such as interactive feedback (e.g. the additional visual effect showing water pouring, played when a virtual jug object is tilted) (as theorised in [229]).

IVR allows for the exploration of this question, as actions in IVR are authored by the designer of the virtual environment, and, unlike actions in the physical world, the fidelity and feedback surrounding an action in a virtual environment is explicitly a design decision.

Hypothesis	**Outcome**
H1. The high feedback group will demonstrate stronger verb learning gains and faster response times than the low feedback group	Rejected
H2. The high feedback group will demonstrate higher presence scores, particularly on the involvement and realism subscales	Rejected
H3. The high feedback group will demonstrate higher value/usefulness scores	Rejected
H4. The high feedback group will demonstrate higher interest/enjoyment scores	Rejected
H5. Direct questions about realism and interactivity will be able to reflect the distinction in the experimental conditions	Not rejected

Table 4.12: Summary table of hypotheses and results

HBSI is one of the most prominent interactions in IVR. However, there are further aspects in which the virtual world attempts to mimic interactive behaviour from the physical world (henceforth referred to as sensorimotor contingencies [295]). For example, haptic feedback, in which the user feels a virtual object's weight via gloves and motors, has been shown to have some impact on users' IVR experiences, promoting faster interaction completion times [33] and providing benefits for object-manipulation tasks [170]. These haptic systems are mostly out-of-reach of consumer-grade IVR systems, however.

A more common and accessible sensorimotor contingency that offers additional multi-modal feedback is the virtualisation of 'realistic' audiovisual feedback from an object interaction. For example, imagine pouring a jug in a sensorimotor-enabled IVR application. Reaching and grabbing a 3D model of a jug and tilting it spout-downwards can be considered an interaction using the hand-based sensorimotor contingency. However, this interaction can be augmented with audiovisual feedback: as the jug reaches a certain level of tilt, a water-pouring particle effect appears, and the sound of running water plays.

While research has begun to elucidate the benefits of the hand-based haptic sensorimotor contingency, this type of 'audiovisual action-feedback' contingency is less-well explored. However, understanding the latter may even be more important, as while high-quality haptic feedback hand-based interactions are currently infeasible for most systems, audiovisual feedback is a design choice in every IVR experience that uses hand-based interactions.

In addition to potentially offering an explanation for what aspect of actions in IVR is causing the additional learning efficacy, exploring this area can also challenge conceptions around best-practice design for IVR systems. Providing rich audiovisual feedback during and after an interaction is often assumed to create a better user experience; often based upon the reasoning that more sensorimotor contingencies are better. However, there is little evidence that this is actually the case, nor whether it is true across different purposes of IVR activity. The sensorimotor contingencies that provide an enjoyable and effective user experience for one IVR use-case may be distinct from another, and both may be distinct from a physical-world experience.

In order to examine whether the above conceptions are true, this section documents an experiment that explores the impact of rich audiovisual action feedback on learners in IVR. Participants either experience a 'low feedback' condition, in which virtual objects can be manipulated (e.g. participants an move a jug); or a 'high feedback' condition in which there is additional audiovisual feedback for object interactions (e.g. the jug displays falling water when poured and a pouring sound is played).

Higher learning gain or faster response times for the high feedback condition would provide evidence supporting the assumption that rich audiovisual feedback enhances user experiences (in terms of user outcomes), and the idea that IVR experiences should follow typical physical world interaction experiences. This result would also support the argument that the enhanced learning gain from actions stems from increased multi-modal feedback, and that increased interactivity and modes of feedback lead to stronger learning outcomes.

However, if no difference is found, then it may evidence that there is a limit to the benefits of replicating the physical world, and the unfettered requirement for many types of sensorimotor contingency. This result leads to an interesting further question - if replicating the physical world is not the ideal goal for HBSI learning in IVR, then what is?

Finding no difference would also provide weight to the argument that learning gains from actions in IVR stem from different cognitive pathways for actions then from gestures.

This study also examines other questions related to this audiovisual action feedback contingency: does high feedback enhance a user's sense of presence in IVR? Does it impact their motivation; or beliefs around how useful or interesting the system is for learning? And what is the meaningful terminology for qualitatively asking users about their experience of this type of action feedback in IVR?

4.3.2 Literature

The literature on how different types of sensorimotor contingency, particularly relating to interactive feedback, impact IVR users' experience is limited. There is initial evidence that the sensorimotor contingency of users using their hands to manipulate virtual objects has positive outcomes for their overall experience and learning outcome (see Chapter 3.1). Additionally, there is growing research that sensorimotor contingencies built on-top of hand-based interaction, such as hardware-based haptics, has an impact on IVR experiences [33, 170, 347]. However, there is little literature regarding the impact of software-based sensorimotor contingencies built on-top of hand-based interactions, such as audiovisual feedback from object manipulations.

Therefore in this section, the available research regarding hand-based sensorimotor contingencies, including haptics, on IVR experience; and present literature justifying language learning as a suitable method for exploring the impact of hand-based, audiovisual interaction feedback on learners in IVR, is discussed.

4.3.3 Hand-based sensorimotor contingency

Increasingly, studies have shown that being able to manipulate virtual objects using your hands or an embodied controller effects users' IVR experiences. The scoping review in this thesis (see Chapter 3.1) found that all of the papers exploring hand-based interaction suggested that hand-base interaction had a desirable impact on user learning outcomes or motivation, and sometimes both. This result was found across various learning topics, including language acquisition, biology, mathematics physics, computer science and geography. Many of these studies specifically involved a direct quasi-experimental comparison between hand-based virtual manipulation of objects and not. However, in such an emerging research area, researchers should be wary that this review may reflect a publication bias in which only experiments with positive results were published and thus analysed.

Beyond learning, there is also evidence that the use of the hand-based sensorimotor contingency is also proving impactful in IVR skill development. Kahlert et al. found that participants could be taught to juggle in the physical world after training for 27 minutes in IVR with hand-based interaction [147]. Rao et al. found that training in a marksmanship task was accompanied by "gradual, robust enhancement of ballistic action and concurrent diminishment of refinements that are likely feedback-moderated" [252]. Reneker et al. found, in a setup enabling a foot-based sensory contingency, "evidence of training and positive transfer from virtual to real-world environments" for soccer athletes [263]. There are limitations to these results, however. In a scoping review, Jensen found that “in general efficient psychomotor skills acquisition with HMDs will not be possible until there are significantly improved peripheral technologies for including the user’s body movements into the simulation” [138]. To reframe Jensen’s comments in Slater’s terminology; the sensorimotor contingency requirements are not being met for certain skill-acquisition applications.

There is also evidence that the use of hand-based sensorimotor contingencies in IVR can influence other cognitive experiences, such as increasing user empathy [57], although in this example it is far more than just the hand-based sensorimotor contingency that varies between the experimental conditions.

Alongside the visual feedback from hand-based manipulations (e.g. seeing manipulated virtual objects move), there has been some research into tactile sensorimotor contingencies, often involving haptic hardware that physically stimulates a user’s hand using vibrations or motors.

There is growing evidence for a relationship between haptic interaction fidelity and user performance, with haptics considered a beneficial additional for surgical training [347], allowing faster task completion times than a non-haptic control [33], and providing benefits in an object-stacking task [170]. Further research explores types of haptic feedback, with “force feedback”, in which a user’s fingers feel resistance inside the IVR system when grabbing or manipulating an object, out-performing vibration feedback in the stacking test (but offering no difference in an object identification test) [170].

Despite research into the impact of the hand-based sensorimotor contingency and the tactile sensorimotor contingency from hand-based interaction, there has been limited research into an interactive audiovisual sensorimotor contingency stemming from hand-based interaction. This is a notable gap, as this type of audiovisual feedback from interaction is a design decision in every IVR system, and one that is not dependent on additional haptic-inducing hardware.

This between-subject experiment examines if there is a distinction between encoding verbs in an environment with high interactional feedback environment (movable 3D objects with congruent audiovisual effects), and in an environment with low interactional feedback (movable 3D objects only). For the high feedback condition, participants used objects to complete actions in order to learn a congruent verb. When an action was completed, audiovisual feedback congruent with the object and the action was provided. For example, a participant had to pick up a jug model and tilt it while learning the Japanese word for ‘pour’. When the jug model was tilted to a certain level, particles representing water would flow from the jug’s spout and the sound of pouring water was played.

In the low feedback condition, participants used objects to complete actions but no audiovisual feedback was triggered (i.e. the jug did not display either audio or visual water pouring effects). In both conditions, participants were given abstract visual feedback that the action had been completed successfully, in the form of a large green tick which appeared when the action was completed.

The learning gain of each condition was monitored and compared to understand if the audiovisual feedback from the hand-based interaction impacted the memorisation of verbs and therefore learning. To investigate potential correlations between feedback condition and presence and motivation, participants in each condition were also surveyed regarding their experience. The relationship between participant presence and motivation and learning gain was also explored. Finally, questions regarding the interaction types were presented in an attempt to understand a method for querying participant experiences regarding this type of system interactivity and feedback.

Hypotheses

The main hypothesis are based on interactionist and multi-modal perspectives that suggest that encoding with high feedback should provide better learning outcomes due to the increased feeling of interactivity and/or additional modes of feedback. Additional hypotheses explore the idea that the feedback conditions may present changes in participants' experience of embodiment, presence and motivation:

- H1. The high feedback group will demonstrate stronger verb learning gains and faster response times than the low feedback group
- H2. The high feedback group will demonstrate higher presence scores, particularly on the involvement and realism subscales
- H3. The high feedback group will demonstrate higher value/usefulness scores
- H4. The high feedback group will demonstrate higher interest/enjoyment scores
- H5. Direct questions about realism and interactivity will be able to reflect the distinction in the experimental conditions

Procedure

Participants were asked to download and run an executable file on their existing IVR systems. An on-boarding process, pre-test, learning process and post-test all took place within the downloaded software.

The on-boarding process explained the IVR control methods, and required users to move to a target location to continue the experience. A voice-over explained the experiment goals and process. The on-boarding process also gave an interactive tutorial of how the learning process worked before launching a pre-test.

Participants were pre-tested for their knowledge of 15 Japanese verbs. The pre-test involved listening to a Japanese word and choosing its English meaning from a list of 15 verbs, or skipping the question. Questions were presented sequentially and participants were not allowed to amend previous answers.

Participants were assigned to one of two interaction groups: high feedback or low feedback. They differ as follows:

- High feedback: Input is made by grabbing the actual VR object, and doing an action that corresponds to the verb being learned. A correct action is given audiovisual feedback congruent to the action, and abstract feedback in the form of a large green tick
- Low feedback: Input is made by grabbing the actual VR object, and doing an action that corresponds to the verb being learned. A correct action is only given abstract feedback in the form of a large green tick

Before the learning process began, participants were asked to practice each action once without any language learning. Each action related to a different object presented in front of the participant on a podium in the IVR. Only objects related to the current action were present at one time. Participants progressed to the next action by grabbing the 3D model and completing the correct action, instructions for which were described aurally. Participants were automatically moved onto the next action three seconds after the previous action was successfully completed. During the experiment, participants were asked to memorise 15 verbs for these actions (see Table 4.13 for list). Participants were exposed to each verb in sequence for four sequences. Participants were told an action, the English verb, and the Japanese language verb. For example, a phrase used for learning 'drink' was "drink from the cup. Drink is nomu. Nomu. Nomu."

For each verb, the participant had to complete the action once and say the verb aloud once. Participants were instructed to say the verb aloud in order to control for the Production Effect, in which speaking a word while encoding it causes stronger memorisation than not speaking it [197]. Both groups of participants used the same avatar - a set of white hands with no arms or body.

After the encoding process, participants repeated the pre-test procedure. For each participant, data relating to correct answers given in their pre-test was excluded, so their learning gain was calculated as the final test result of their eligible question set. Participants who had a large existing knowledge of the target language (above 30% correct in the pre-test) would have been removed (no participants had this level of pre-existing knowledge).

The data collection was done inside a VR environment. To verify the actions were completed correctly, telemetry of the participant's movements was recorded, as in the previous experiment.

A further web-browser-based test was taken one week after the initial study to determine their retention of the information.

Verb	**Action**	**System feedback**
Weigh	Pick up a cat and place on scales	When grabbed, cat meows. When placed on scale, makes a beep and scale display shows a weight
Swing	Pick up a baseball bat and swing it into a baseball	When bat touches baseball, collision sound is played and ball moves relative to collision force
Write	Pick up a pen and write on a notepad	Pen leaves an ink trail on the notepad and a sound of a pen scratching on paper is played
Push	Push a button downwards	A button click is played and a trapdoor opens and drops an object
Throw	Pick up a dart and throw it into a large target	Dart makes a collision noise when hitting the target and sticks into it
Hammer	Pick up a ball hammer and hammer a nail with it	Hammer makes a collision noise when hitting the nail, nail is moved into a piece of wood when hit
Pour	Pick up a jug and tilt it	When tilted, pouring sound and water particle effect are played
Bowl	Pick up a bowling ball and bowl it into bowling pins	Ball makes collision noise when hitting pins, pins are moved/knocked over in the collision
Stir	Pick up a spoon and stir a coffee	Liquid stirring sound is played and coffee liquid rotates in mug
Turn	Turn a dial on a radio model	Radio tuning sound is played, and indicator is moved
Paint	Pick up a paintbrush and paint on a canvas	Painting noise is played, paint appears on the canvas
Switch	Touch a lightswitch, which moves inwards	A clicking noise is played and a lightbulb lights up
Strike	Pick up a match and rub it on the side of a matchbox	Matchstrike sound and particle fire effect is played
Bang	Pick up a gong mallet and bang a gong	Gong noise is played and gong moves in its frame
Cut	Pick up a knife and move into bread	Cutting noise is played, bread slice separates from bread

Table 4.13: List of words to learn, the required hand-based sensorimotor interaction, and a description of the additional feedback given for the high feedback condition

Participants

Seventy-four (74) participants responded to advertisement for the study. Of these, fifty-four (54) participants' data was usable in the final analysis. The reason for participant ineligibility was due to having taken part in only one part of the study - either the VR experiment or the post-experiment survey - and not both. No participants were excluded for having high levels of pre-existing Japanese knowledge.

All participants were recruited from an advertisement posted on the Reddit /r/oculus community and used their own IVR hardware in their own setting. This recruitment method suggests the participants were experienced in using IVR hardware.

The average (mean) age of valid participants was 25 ($SD = 7.79$). Participant gender skewed heavily male (50) over non-binary (3) and female (1), or other/did not say (0). Valid participants had a low knowledge of the target learning words during the pre-test, with only 12 participants having any knowledge of the included Japanese verbs. Of these 12, the average knowledge was 1.17 Japanese verbs.

The majority of valid participants were fluent in more than one language (25 reported as fluent in one language, 22 self-reported as fluent in two languages, 5 in three languages, and one in four languages. One participant reported being fluent in zero languages). There was not a significant correlation between languages known and learning outcomes (r = -0.21, $p = 0.12$).

Interaction condition was randomly assigned inside the software once it was downloaded onto a participant's computer. As such, 32 participants were assigned to the action with feedback condition and 22 to the action without feedback condition.

Corpus

Participants were tested on their knowledge of 15 concrete action verbs. Action verbs were chosen as they are highly embodied and were used in previous gesture and action word memorisation comparisons [334]. The target words were chosen to be familiar actions that allowed for mostly distinct gestures for each word. They were also chosen based upon the ability to present obvious visual and aural feedback resulting from an interaction. For example, hammering a nail allowed the feedback condition to visually present the moved nail's position (from pointing slightly in the wood to being completely embedded) and to aurally present a 'bang' sound effect. This second requirement meant there was a difference in verbs from the prior experiment.

Japanese was chosen as a language due to its obscurity for many native

English speakers, and distance [54] from English. Japanese gairaigo (import words) were specifically avoided to reduce the chance of participants' inferring a meaning from their similarity to English. Phonetically similar and particularly long Japanese words were also excluded, as there were concerns that beginner-level learners would find these words difficult to tell apart. A list of these words, the participant's action and the feedback given by the system can be found in Table 4.13

Environment

A 3D environment similar to the previous experiment was created in Unity (2019.2.21), with the Unity XR Interaction Toolkit for handling IVR interaction. The environment was explorable via a head-mounted display and embodied controllers. Navigation could be done by moving around the real space and/or by using the thumbsticks on the controllers.

Evaluation

Participants' knowledge of the verbs was evaluated in the same way as in 4.2. This means their verb knowledge was measured in three tests: one administered before their exposure to the environment (pre-test); one immediately after (post-test), and one seven days later (week-test). Participants performed the same test each time, listening to a Japanese word and choosing the English meaning from a list of 15. All three tests were conducted remotely outside of laboratory conditions; the first two were conducted inside IVR, and the third was via an HTML form on a web browser. The time taken for each question was recorded to help evaluate the testing sessions. A visual examination of this data did not highlight any individual user taking a consistently long or short time to answer each question, suggesting that participants avoided looking-up answers; being consistently distracted (in a way that could be measured by time) during the evaluation; or rapidly entering answers in order to receive payment. Participants were not given feedback when submitting answers.

Learning gain was calculated as a normalised score between 0 and 1, measured as post-test score (excluding words answered correctly in the pre-test) divided by the number of eligible words for their session (eligible words were words they had not got correct in their pre-test):

$$(PostScore - PreScore)/(EligibleWords - PreScore)$$

Twelve participants had existing knowledge of either one or two of the target verbs - these were removed from their pre-test, post-test and eligible words calculations. It was recorded whether participants listened to the audio clip before submitting an answer - this was the case for every entry.

After using the system, participants were asked to complete a survey in-browser. This consisted of the Igroup presence questionnaire [282], the Intrinsic Motivation Inventory intrinsic motivation questionnaire [268]) and three bespoke questions regarding the experience of realism, interactivity and environmental impact.

Bespoke questions were used concerning realism, interactivity and environmental impact in an attempt to understand how users contextualise this type of action-feedback. Participants were asked to rate whether the actions they made while learning were 'not realistic (1)' to 'very realistic (5)'; whether the overall interactivity of the experience was 'not interactive (1)' to 'very interactive (5)'; and whether they felt their actions were 'not having an impact (1)' to 'having a large impact (5)'. The previous experiment used the Gonzalez-Franco & Peck embodiment questionnaire [103], but it was felt that it was unable to detect this type of sensorimotor interaction distinction between interaction groups. This is likely because it is more focused on the feeling of embodiment in a virtual avatar, rather than embodied sensorimotor activity.

The questionnaires used for these can be found in Appendix C.

Cognitive load was again not measured in this experiment, based upon the result from the first which suggested that, if there was a real difference in cognitive load, then the self-reported measure of cognitive load used was not sufficiently sensitive to detect it. As this study was remote, it was infeasible to deploy physiological measures of cognitive load to address this; and there was a desire to limit the number of questions given to the remote participants.

Analysis

The first hypothesis (the high feedback group will demonstrate stronger verb learning gains and faster response times than the low feedback group) was tested by coding correct responses as 1 and incorrect responses as 0. Where a participant had answered correctly in the pre-test, their responses for that word were removed. Mixed Models was used to account for both the fixed (interaction type) and potential random (users, words) effects, as recommended by Macedonia et al. [194]. As the dependent variable was binomial, a Generalised Linear Mixed Model was used.

For the response times, a Linear Mixed Model was used due to the continuous dependent variable of response time. Only correct answers were included in the dataset, and outliers were removed. Four outliers were removed, as the participant took over one minute to respond to the question. It was considered accepted to remove these outliers as some participant's mean answer times were skewed by a few longer entries, potentially caused by distracting out-of-lab circumstances.

For the second hypothesis (the high feedback group will lead to higher presence scores, particularly on the involvement and realism subscales), independent t-tests were calculated for each of the four presence scores calculated from survey results (general, spatial, realism, involvement) between the interaction conditions (low feedback and high feedback).

For the third hypothesis (the high feedback group will demonstrate higher value/usefulness scores), independent t-tests were calculated for each of the four presence scores calculated from survey results (value/usefulness) between the interaction conditions (high feedback and low feedback).

For the fourth hypothesis (the high feedback group will demonstrate higher interest/enjoyment scores) independent t-tests were calculated for each of the four presence scores calculated from survey results (interest/enjoyment) between the interaction conditions (high feedback and low feedback).

For the fifth hypothesis (direct questions about realism and interactivity will be able to reflect the distinction in the experimental conditions), independent t-tests were calculated for each of the three questions: (1) On a scale of 1 - 5, with 1 being not realistic and 5 being very realistic, how realistic would you consider the actions you made while learning? (2) On a scale of 1 - 5, with 1 being not interactive and 5 being very interactive, how would you rate the overall interactivity of the experience? and (3) On a scale of 1 - 5, with 1 being not having an impact and 5 being having a large impact, how much of an impact did you feel your actions had on the environment? These were also between the interaction conditions (low feedback and high feedback).

4.3.4 Results

H1. The high feedback group will demonstrate stronger verb learning gains and faster response times than the low feedback group

A Generalised Linear Mixed Model was used to understand if the conditions had an effect on learning gain (see Table 4.15). For the post-test, the GLMM (n = 826; 54 participants) showed learning gain varied across both participants ($\sigma^2 = 1.43$) and words ($\sigma^2 = 1.50$). After controlling for these random factors, the model did not present a statistically significant distinction in learning gain between words encoded in the high feedback group ($p = .681$, $\beta = 0.15$, 95% CI [-0.59,0.90]) and those in the low feedback group. Although not significant, in the model, given a participant and word with average intercepts, the probability of getting a correct response increases from 36.2% to 39.8% in the feedback group.

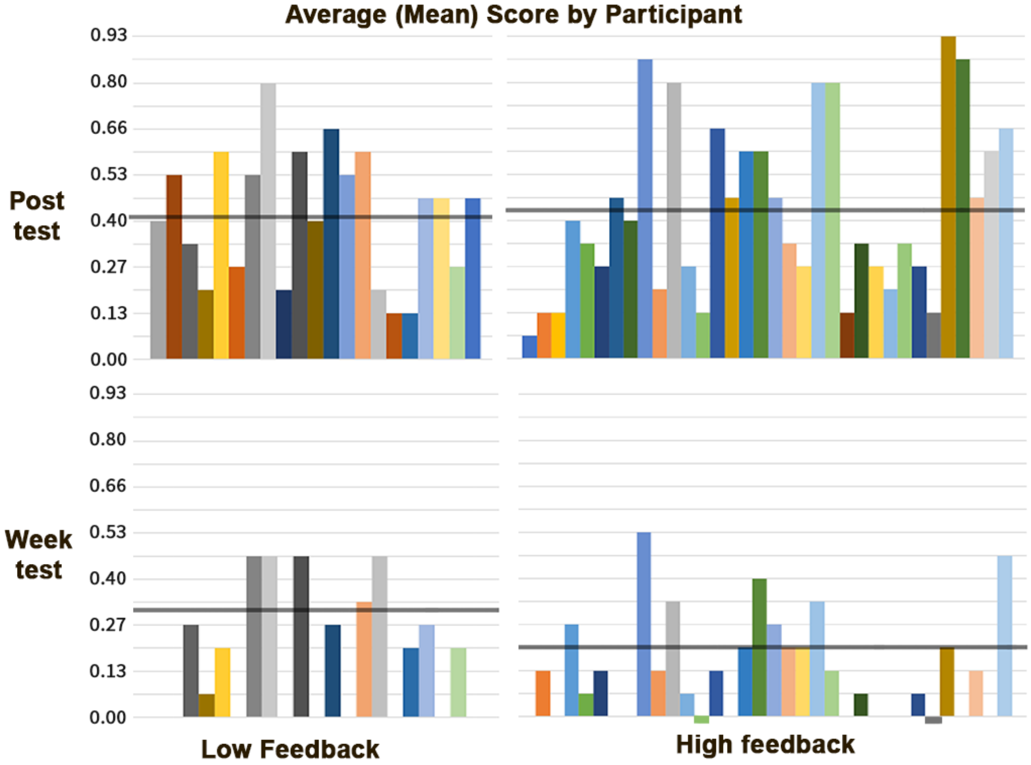


Figure 4.13: *Participant Learning Gain.* Each bar shows the learning gain (number of correct test answers minus existing knowledge) of a participant. Results are displayed for each test, either the one taken immediately after learning (Post-Test) or the test conducted a week later (Week-Test). Participants are grouped by interaction condition (low feedback and high feedback). If a participant did not take part in the Week-Test, a gap is shown; a coloured mark on the zero line represents the Week-Test was taken but no correct answers were given. The mean learning gain of each condition and test is also shown; the learning gain after one week for the low feedback condition was significantly higher

For the week-test, the GLMM (n = 517; 35 participants) also showed learning gain varied across both participants ($\sigma^2 = 0.4$) and words ($\sigma^2 = 0.72$). After controlling for these random factors, the model presented a statistically significant relationship between interaction type and learning gain ($p = .014$). Words encoded in the high feedback group were remembered worse than those in the low feedback group (β = -0.79, 95% CI [-1.48,-0.15]). In the model, given a participant and word with average intercepts, if they were assigned to the low feedback condition, they would be 14% more likely to correctly remember a word than in the high feedback condition (16% vs 30%).

A Linear Mixed Model was used to understand if the conditions had an effect on response time of correct answers (see Table 4.9). For the post-test, the LMM

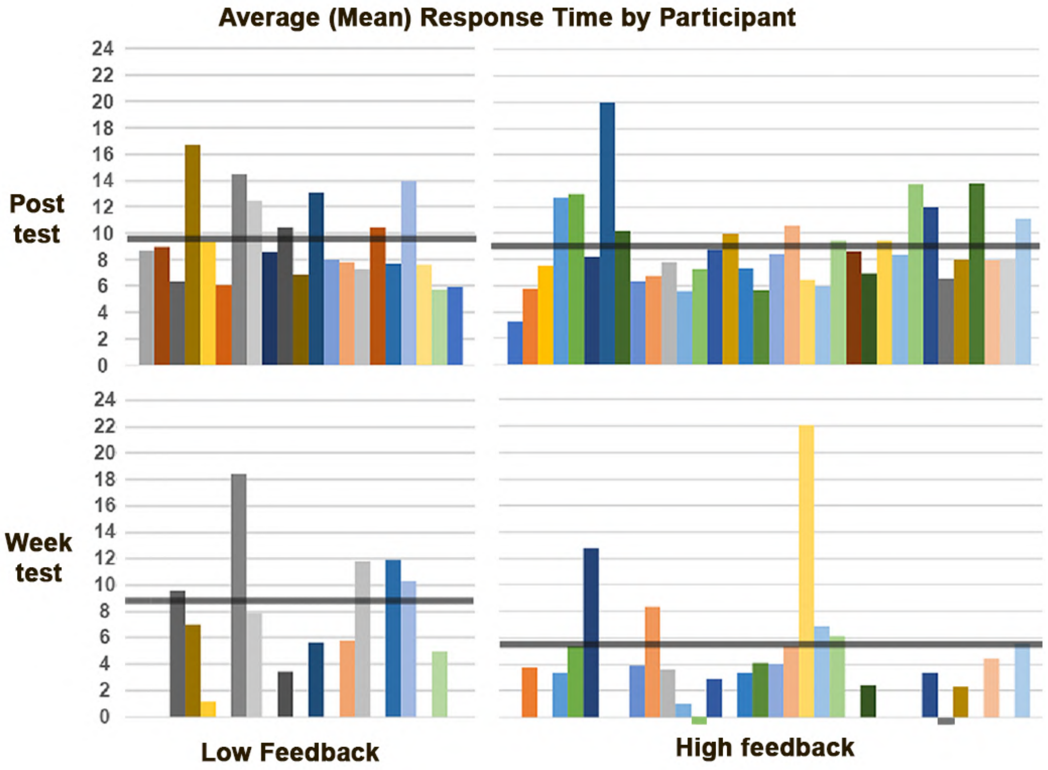


Figure 4.14: *Participant Response Times.* Each bar shows the average response time of a participant's correct answers (in seconds). Response times are displayed for the test immediately after learning (Post-Test) and the test conducted a week later (Week-Test). Participants are grouped by interaction condition (low feedback and high feedback). If a participant did not take part in the Week-Test, a gap is shown; a coloured mark on the zero line represents the Week-Test was taken but no correct answers were given. The mean response time of each condition and test is also shown, which had no significant difference.

(n = 352; 55 participants) showed response time varied across both participants ($\sigma^2 = 6.651$) and words ($\sigma^2 = 4.71$). After controlling for these random factors, there was no evidence of a significant distinction between between the feedback group ($p = 0.554$, β = -0.537, 95% CI [-2.33,1.27]) and the non-feedback group.

For the week-test, the LMM (n = 122; 35 participants) showed response time varied across participants ($\sigma^2 = 8.516$) and to a lesser extent words ($\sigma^2 = 5.724$). After controlling for these random factors, there was no evidence of a significant distinction between between the high feedback group ($p = 0.931$, β = -0.541, 95% CI [-5.41,0.91]) and the low feedback group.

Descriptive statistics for learning gain and response times by feedback condition and test type (post or week) can be found in Table 4.14. To see a graph of the learning gain and total response time of each participant by feedback condition and test type, check Fig. 4.13 and Fig. 4.14 respectively.

As there was a significant result of a small effect size in the opposite direction of this hypothesis, but only in the retention test, the hypothesis is not accepted.

Results	**N**	**Mean Score**	**Mean RT**
Low: Post-test	22	0.41 ±0.25	9.68 ±3.73
High: Post-test	32	0.44 ±0.25	9.14 ±3.28
Low: Week-test	12	0.32 ±0.23	8.78 ±4.09
High: Week-test	23	0.20 ±0.20	5.46 ±2.79

Table 4.14: Table of learning gain results (normalised mean) and response time (mean) with standard deviation from tests immediately after the learning session (post-test) and one week later (week-test)

Parameter	**Beta**	**Lower-95**	**Upper-95**	**SE**
		Learning Gain Post-Test (GLMM)		
Intercept	-0.57	-1.47	0.29	0.43
Interaction	0.15	-0.59	0.90	0.37
		Learning Gain Week-Test (GLMM)		
Intercept	-0.87	-1.59	-0.20	0.38
Interaction	-0.79	-1.48	-0.15	0.32
		Response Time Post-Test (LMM)		
Intercept	9.81	7.96	11.67	0.92
Interaction	-0.54	-2.33	1.27	0.90
		Response Time Week-Test (LMM)		
Intercept	8.61	5.87	11.36	1.36
Interaction	-2.28	-5.41	0.91	1.56

Table 4.15: Table of Generalised Linear Mixed Model results for learning gain and Linear Mixed Model results for response time. Interaction refers to high feedback condition. Note: GLMM co-efficients are logit

H2. The high feedback group will lead to higher presence scores, particularly on the involvement and realism subscales

Presence Type	**Feedback**	**No Feedback**	**P**
General	1.19 (±1.33)	1.36 (±1.30)	.64
Spatial	1.03 (±0.85)	1.02 (±0.92)	.99
Involvement	0.01 (±0.82)	0.00 (±0.72)	.97
Realism	-0.42 (±0.61)	-0.35 (±0.57)	.69

Table 4.16: Summary of mean presence scores and the significance of their relationship with feedback type, either with feedback or with no feedback, with standard deviation

In the independent t-tests to explore the relationship between interaction group and presence (and its subscales), no significant effects were found. These (and all later t-tests) were found to satisfy equality of variance conditions (Levene's test). A summary of these results can be found in Table 4.16.

There was no significant effect for general presence (95% CI [-0.57, 0.92], $t(52)$ = 0.473, p = .638) between feedback (M = 1.19, SD = 1.33) and non-feedback (M = 1.36, SD = 1.30).

There was no significant effect for spatial presence (95% CI [-0.50, 0.49], $t(52)$ = -0.016 p = .987) between feedback (M = 1.03, SD = 0.85) and non-feedback (M = 1.02, SD = 0.92).

There was no significant effect for involvement presence (95% CI [-0.45, 0.43], $t(52)$ = -0.036 p = .972) between feedback (M = 0.01, SD = 0.82) and non-feedback (M = 0.00, SD = 0.72).

There was no significant effect for realism presence (95% CI [-0.27, 0.40], $t(52)$ = 0.407 p = .685) between feedback (M = -0.42, SD = 0.61) and non-feedback (M = -0.35, SD = 0.57).

As no significant relationship on any of the presence scales were found, H2 is not accepted.

H3. The high feedback group will demonstrate higher value/usefulness scores

The independent t-test between interaction group and value/usefulness found a significant difference (95% CI [0.00, 1.37], $t(52)$ = 2.030 p = .047) between high feedback (M = 5.10, SD = 1.37) and low feedback (M = 5.79, SD = 0.90), in the opposite direction than expected. It had a moderate effect size (d = 0.595).

Therefore H3 is rejected.

H4. The high feedback group will demonstrate higher interest/enjoyment scores

The independent t-test between interaction group and interest/enjoyment did not demonstrate a significant difference (95% CI [-0.38, 0.93], $t(52) = 0.854$ $p = .397$) between high feedback ($M = 5.85$, $SD = 1.31$) and low feedback ($M = 6.13$, $SD = 0.89$).

Therefore H4 is not accepted.

H5. Direct questions about realism and interactivity will be able to reflect the distinction in the experimental conditions

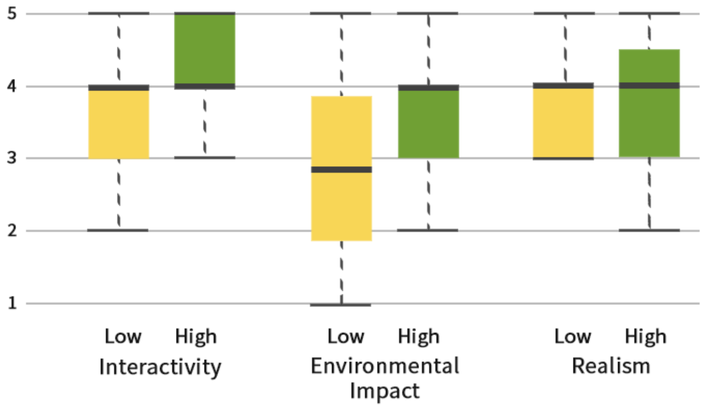


Figure 4.15: Responses to direction questions about the low feedback and high feedback's interactivity, environmental impact and realism. Only 'interactivity' showed a significant difference between experiences, with the high feedback considered more interactive

The independent t-test between interaction group and perception of interactivity question demonstrated a significant difference (95% CI[-0.90, -0.03], $t(52) = -2.14$ $p = .037$) between high feedback ($M = 4.28$, $SD = 0.76$) and low feedback ($M = 3.82$, $SD = 0.78$). The effect size was large ($d = 0.85$).

The independent t-test between interaction group and perception of environmental impact question did not demonstrate a significant difference (95% CI[-1.13, 0.11], $t(52) = -1.66$ $p = .102$) between high feedback ($M = 3.37$, $SD = 1.08$) and low feedback ($M = 2.86$, $SD = 1.10$).

The independent t-test between interaction group and perception of realism question did not demonstrate a significant difference (95% CI[-0.87, 0.34], $t(52)$

= -0.874 p = .386) between high feedback (M = 3.72, SD = 1.01) and low feedback (M = 3.45, SD = 1.16).

A box plot of these results can be seen in Fig. 4.15.

As one of these questions demonstrated a significant difference between the groups, with a large effect size, H5 is accepted.

4.3.5 Discussion

Rich feedback does not mean better learning

The experiment did not show that the high feedback condition provided additional learning gain or faster response time compared with the low feedback condition; either for immediate or long-term memorisation. This lack of benefit was not expected, and could be used to argue for limits to theories of learning based on the idea that more modes of feedback lead to better learning outcomes, such as the memory trace theory [17]. It also questions whether the creation of “first-order” IVR is truly useful for learning in IVR, and whether this could be the case for other IVR use-cases.

The fact that this experiment found that the low feedback condition had significantly stronger long-term learning gains could suggest that the richer feedback actually harms learning. If this result is put in the context of the previous experiments, which have found that encoding while doing hand-based sensorimotor activity in IVR caused stronger memorisation than no sensorimotor action, and that hand-based interaction with objects in IVR caused strong memorisation (actions) than hand-based activity that did not interact with objects (gesture), then it is possible to suggest some kind of interaction-learning model in VR (at least for verb memorisation). In this model, the benefits of hand-based sensorimotor learning in IVR peak with hand-based sensorimotor interaction that involves manipulating a virtual object, but that additional audiovisual feedback related to the manipulations is at best, not helpful, and at worst, harmful.

Interestingly, the results from the intrinsic motivation survey also reflect the learning outcomes: participants reported that the low feedback condition was more usefulness and had more value to them, with a notable effect size. However, in the free-form feedback survey at the end of the study, no participant stated that the high feedback was distracting or that it was directly harmful to their learning, so it is unclear why participants (on average) felt this way, or how this result could relate to the long-term learning results.

One possible answer could be that the rich feedback increases a learner’s cognitive load, harming their retention. There is some evidence that cognitive load may be a particularly sensitive issues for learning in IVR, as well as a key contributing factor to IVR learning outcomes [203]. It would be useful for a future experiment to explore whether cognitive load was a factor in these results, or whether the model suggested above, in which hand-based sensorimotor learning benefits in IVR peak with hand-based sensorimotor interaction that involves manipulating a virtual object, was the explanation.

An alternative explanation is that the high feedback condition in this experimental was poorly realised, and it was the frustrating or negative experience that had a detrimental effect on learning. It is hard to evidence this argument without a quality of interaction or feedback rating for the experience. However, if participants' interest and enjoyment are used as a proxy for understanding how well the experience was realised, then there was no significant distinction between the groups, and therefore no evidence for this theory.

Another possibility is that the high feedback was distracting, as it would cause participants to 'play' with the objects in order to repeatedly experience the feedback. However, given that (a) these were non-novel VR users (using their own hardware), (b) that each interaction had to be repeated multiple times (and so may be less interesting to 'play' with on the second and beyond repetitions), and (c) that many of the actions could only be conducted once per trial (i.e. the nail did not remove itself from the wood after being hammered); this would likely have had a limited impact on outcomes.

Presence: not the IVR learning panacea

The distinction in long-term learning gain between the two conditions was not reflected in any of the presence measurements, which adds further evidence refuting the idea that presence is the primary driver of IVR-based learning, especially for IVR sensorimotor-based learning.

This result also suggests that rich feedback does not impact any type of presence a participant experiences. This finding is somewhat unexpected, given that one might expect place illusion and plausibility [295] to benefit from a more 'realistic' simulation with higher sensorimotor contingencies.

One explanation for the lack of a distinction in presence ratings between the conditions could be that rich feedback for sensorimotor interactions has a very minor impact on a user's presence experience compared with more prominent experiential design choices. For example, the abstract setting and activities could have been far more influential on how users experienced presence. Further evidence for this explanation comes from that lack of significant difference in how users rated the 'realism' of the experience, when it one would imagine that the additional audiovisual feedback is more 'realistic' than it not being there.

Questions for understanding action feedback

The bespoke sensorimotor and embodied survey questions presented some evidence of how participants understand the distinction in feedback conditions.

There was a significant difference between the groups' responses to the question "how would you rate the overall interactivity of the experience?", with a large effect size, but no significant difference for responses to 'realism' or 'impact on the game environment'. This suggests that, for understanding this type of feedback distinction in conditions, realism or impact on environment may not be useful terms for enquiring with users.

Participants considered the high feedback condition more 'interactive', which suggests that users considered the audiovisual feedback of an action a contributor to interactivity. This is true despite the core gameplay interaction already providing abstract feedback for both conditions (a green tick when an action was correctly enacted and stage advancement to the next word).

This result could suggest an important distinction between 'agency' and 'interactivity'. The agency in both conditions was the same (participants could manipulate virtual objects), but the 'interactivity' was considered different. Given that the more 'interactive' condition did not create a stronger sense of presence or result in a better learning outcome, this could suggest that researchers should treat agency and interactivity more distinctly, particularly for IVR sensorimotor activities.

4.3.6 Limitations

One of the limitations of these results is that a notable number of participants did not take part in the follow-up test, which reduced the sample size from 56 to 35. This had a particular impact on the response time linear mixed model, where only correct answers were included, and so the sample narrowed both due to participants forgetting answers over time and drop-out rates. However, it is expected that the smaller sample would make it less likely for the data to present a significant distinction between the groups in the retention test, and the results demonstrated the opposite regarding learning gain. Still, it would have been better if participants had been retained to a greater extent.

Another notable limitation is the non-quantified quality of the high feedback. It is impossible to empirically argue that the feedback condition has implications for learning when the quality of feedback was not quantified. Broken, frustrating, unclear and irrelevant feedback could all harm the learning experience, in a similar fashion to how distraction, frustration and in-congruence can harm all learning. There is limited evidence here to suggest that the types of high feedback included in the study avoided these pitfalls. However, qualitative feedback from three participants suggested that only the “pen and writing” high feedback experience was notably problematic, describing it as “jittery”. However, as this action was criticised and no others, it is reasonable to assume the other high feedback tasks were unproblematic.

Finally, this experiment did not attempt to investigate whether different player types had different performances depending on the feedback type they experienced. For example, it could be that a participant’s learning modality preference might have impacted the results [85], and that there are types of players who greatly value high audiovisual interactional feedback. While this aspect is not explored here, it is certainly a consideration for further work.

4.3.7 Conclusion

This experiment provides evidence opposing the idea that richer feedback in IVR will improve learning and retention. In fact, it presents evidence that higher action feedback in IVR learning could actually harm learning outcomes, evidenced by a higher learning gain after one-week for the low action feedback condition. This opposes the idea that the ultimate goal of IVR should be to create a system in which the traditional, physical world experience is virtualised. The results also counter theories of learning and memorisation that suggest more modes of interactional feedback created stronger learning outcomes.

The negative relationship between action feedback and retention encourages a nuanced approach to learning system design, and opposes a ‘more-real-is-better’ approach. How to optimise learning in IVR when the ideal is not in matching reality, and indeed, more closely matching reality harms learning outcomes, is a fascinating challenge, but also a fascinating opportunity. Perhaps this result, and others like it, can start to move us away from the perspective that the physical-world-is-best, and into exciting new spaces in which there are interesting, unique optimisations for learning outcome that come from learning in virtual realities as opposed to physical ones.

Chapter 5

Discussion and Conclusions

This thesis has presented evidence that if, and how, sensorimotor interactions are implemented in IVR can have a notable impact on users' cognitive outcomes (Chapter 5.1.1). It provides evidence that there are similarities between sensorimotor learning in IVR and the physical world, in terms of types of sensorimotor encoding and their outcome; and that sensorimotor interactions taken in IVR are, if not entirely the same as physical world actions, then at least different from physical world gestures, even though the object of the IVR interactions is entirely virtual. It also provides evidence that rich interactional feedback provides a distinction in perceived 'interactivity' by participants, but that the distinction does not cause a positive cognitive effect (in terms of learning).

Finally, it speculates on the theories of cognition that could explain the results (Chapter 5.1.3), arguing that there are distinct action and gesture encoding pathways, and that the action pathway benefits from a sense of agency and achievement of desired interactions, which enable the observed benefits of actions for learning; and that the learning benefit of sensorimotor interaction is not based upon the increased number of sensory modalities.

In the experiments, participants' second language verb learning was the cognitive outcome measured; a topic that is considered highly embodied and has strong roots in sensorimotor-enabled learning. Therefore this thesis also provides evidence that learning verbs through IVR sensorimotor interaction is an effective approach for computer-aided language learning (Chapter 5.1.2).

Although not the predominant concern of the research, the impact on presence of how sensorimotor interaction is incorporated in IVR is also discussed (Chapter 5.1.4).

5.1 Discussion

The full details of the relationship between human cognition and sensorimotor activity is currently unknown. The virtualisation of human sensorimotor interactions into IVR is similarly mysterious: it both broadens our sensorimotor possibilities into potentially impossible activities and environments; and constrains or changes them through the artefacts and limitations of the immersive hardware used and the artificial environments deployed.

Part of the difficulty in understanding and examining sensorimotor interaction and cognition in IVR is the breadth of the possibility spaces; the experiments deployed in this research are hugely artificially constrained explorations containing a myriad of design choices and potential influential dimensions, informed by best-guess and best-practice. These factors could have potentially profound implications that are not yet understood, as it is not just that the influences and impacts of IVR factors are unknown, but also the factors of IVR that could influence and impact.

This thesis is bound to two major assumptions regarding the experimental paradigms used. The first is that the experiments are accurately able to explore sensorimotor cognition without the experimental process obfuscating it. This is an assumption on which much cognitive science research is based, but it is potentially more crucial in IVR where the entire interactive environment is artificially authored. The second is that, by comparing the outcomes of controlled experiments, more specifically, the learning gain outcomes between different interaction conditions, we can intuit the relationships between sensorimotor conditions and 'cognition'.

Based upon the second assumption being broadly correct, this chapter discusses the findings of this thesis in four aspects: the findings' implications on conceptions concerning sensorimotor interaction in IVR, and on IVR sensorimotor interaction affordance (Chapter 5.1.1); the implications and limitations of this research on IVR language learning (Chapter 5.1.2); the wider theories of embodied learning and memorisation that our results evidence (Chapter 5.1.4); and the implications on other factors related to IVR, such as presence and motivation (Chapter 5.1.3).

5.1.1 Sensorimotor interaction in IVR and cognition

The first step in understanding the impact of sensorimotor interaction in IVR is establishing whether the use of sensorimotor interaction by users provides any distinction over non-sensorimotor interaction.

This was evidenced in the first experiment, which demonstrated a significant and notable learning gain for learners who used sensorimotor interaction in their learning compared with those who did not. Therefore, sensorimotor interaction can (and did) have an impact on cognition (learning) in IVR.

Although this experiment established that a user who used sensorimotor interactions in IVR was likely to learn more than one who did not manipulate their body at all (besides looking around with their head), this result was ambiguous as to what aspect of the sensorimotor activity caused the distinction in learning. Language learning literature overwhelming suggests that bodily activation helps with language memorisation. Therefore, a preeminent question is whether the cognitive distinctions are caused only by the different levels of body activation, or whether the interactivity provided by the IVR had a role in the outcomes.

The second experiment demonstrated that the cognitive impact of sensorimotor interaction in IVR was dependent on the virtual interactions that a participant could take, and not just on bodily activation. Participants who manipulated congruent virtual objects while learning, learned more than those who manipulated their bodies in a similar way but without interacting with the virtual objects. Therefore, the observed cognitive impact stems not from the fact that the body is being moved, but that the body is both moving and interacting with the virtual environment.

These findings replicate results found in similar physical world investigations in which actions provided slightly stronger encoding than gestures. In the physical world, gestures and actions are often considered as distinct (but linked) cognitive processes, with different cognitive frameworks linked to each type. The distinction discussed above, between body manipulation and interaction in IVR, suggests that the sensorimotor interaction afforded by IVR is rich enough to create a distinct encoding experience between user's gestures and actions.

Therefore, an action-based framework is likely more suitable for understanding our object-based interactions in IVR, while a gesture-based framework is also suitable for gestures (e.g. waving). This means researchers should avoid the idea that because a user is not physically touching a physical object, that the user's interactions should always be treated as gestures or approached from within a gesture-oriented framework. It might also be worth considering the retirement of terms such as 'gesture controllers', given that there is now evidence that those controllers enabled, as far as learning is concerned, both gestures and actions.

Having established that the manipulation of objects has a cognitive impact on IVR users, the next step was to understand if there were observable variances caused by feedback from the object manipulations. Is the cognitive difference caused by our ability to manipulate a virtual object, or is it because we manipulate the object and the virtual word provides greater context around that interaction, such as audiovisual feedback?

The results in the third experiment showed similar learning happens whether the audiovisual feedback from interacting with an object in IVR is rich or not, with worse long-term retention results for sensorimotor interaction that offered high levels of audiovisual feedback in response to manipulations. One reading of this is that, for an 'action' to occur in IVR, users only need to be able to manipulate a virtual representation of that object, rather than for it to behave either in a way similar to the physical world, or for it to have rich feedback.

Together, these results can be used to suggest a model of sensorimotor interaction in IVR and its impact on learning (and potentially cognition generally). In this model, sensorimotor-based object manipulations in IVR create a distinctly different cognitive outcome than no sensorimotor activity, or sensorimotor activity that does not involve interacting with an object. However, beyond the manipulation of a 3D representation of an object in virtual space, further feedback provides no further benefit. This suggests that it is the agency offered by sensorimotor interaction - the ability to interact and manipulate an object in a way that completes a given task - rather than the multi-modal richness, that creates a notable cognitive impact.

With this said, it is important to note that there is almost certainly no singular rule that outlines the impact that sensorimotor interaction has on cognition in IVR. Research into memory and learning, as explored here, may observe a strong impact from sensorimotor interaction in IVR, whereas research in empathy or guilt may find none at all.

Similarly, it is also highly unlikely that there is a singular rule that virtual sensorimotor interactions produce similar cognitive outcomes to physical sensorimotor interactions. For example, it is unlikely that swinging a virtual sword to kill a virtual enemy would have the same impact on a user as if doing so in the physical world, with a physical sword.

It is more likely that the cognitive impacts of sensorimotor activity in IVR, and their relationship to physical world sensorimotor activity, is heavily dependent on the subject matter. In this thesis, we explored learning and memorisation, and therefore we should be cautious about extrapolating any findings too far towards a universal rule of IVR sensorimotor interaction.

5.1.2 IVR language learning

At the most specific, the results of the experiments in this thesis could be said to apply only to cognition concerning concrete action verb learning in one-off interaction settings, for the purpose of training learners to aurally recognise second language words via listening tests.

In this well-defined area, the claims of this research can be fairly bold: actions were better than no actions for immediate word retention; that actions were better than gestures for immediate word retention; and no difference in immediate word retention between actions with or without feedback were found.

These findings could be used as design suggestions for the creation of IVR language learning software (given an assumption that learning to aurally recognise words is indicative of quality second language education). To enhance learning efficacy, the results here suggest that designers should ensure that concrete action verbs are taught in a way that enables sensorimotor interaction. This should be done through congruent verb actions which allow the learner to manipulate objects related to the action. However, additional feedback beyond object manipulation is less important, and potentially detrimental to the learning experience. For example, it is important we allow learners to swing a golf club to learn the verb 'swing', but less useful to create a ball simulation to handle a resulting impact between the club and ball.

There is, however, the results have a notable absence of evidence regarding longer-term retention between the sensorimotor conditions. The results presented here showed no significant evidence that a single sensorimotor-enabled learning intervention provided more learning than a non-sensorimotor-enabled one. However, it is intuitive to say that immediate learning gain should turn into longer-term retention through regular learning sessions; and it is possible to explain away this lack of retention through smaller margins of difference in learning gain (due to general forgetting) and smaller sample sizes due to experiment drop-out. In all, given the strength of the learning gain distinction when tested immediately, it seems unlikely that this would not carry over to retention benefits if sensorimotor-enabled learning was used regularly.

It is interesting to consider how sensorimotor-learning might aid non-concrete verb learning. Evidence from the first experiment did not find a significant difference between interaction conditions for noun learning, but there was a distinction in mean, and average learning was smaller in general for nouns than for verbs, making an effect harder to detect. Considering the results from the second and third experiments, in which the conclusion is that object manipulation, not body activation nor high-level object interaction feedback, is the potent driver of learning, then it is logical to suggest that noun manipulation would result in a similar benefit as verb actions. The lack of significant evidence for this is disappointing, and it would be interesting to see further, more powerful studies investigating this.

Additionally, the sensorimotor interaction impact on adjectives, adverbs, tenses, communicative competence and many more are all missing explorations in any comprehensive understanding of how sensorimotor-interaction might impact a holistic IVR language learning approach.

5.1.3 Embodied, sensorimotor learning and memorisation

Researchers have been wrestling with why sensorimotor activity leads to better memorisation inside of embodied cognition framework. The two predominant theories are (1) activation of the body causes strong memorisation; or (2) multi-modality causes memorisation, with body-based interaction proving richer modalities than more abstracted forms. Proponents of the first theory argue that memorisation is primarily enhanced because learners are engaging their bodies; while for the latter the body is just one of many potential ways of adding to the multi-modality, or adding richer 'memory traces' to a 'learning episode'.

This thesis refutes an absolutionist view of either perspective. The results in the second experiment show that activity that featured body activation with object manipulation, rather than limited to bodily activation alone, provided better learning outcomes, which is not consistent with a body-only perspective. However, the results in the third experiment, demonstrated that additional feedback from the interactions (in the form of audiovisual feedback as objects were manipulated) provided no advantage over object manipulations without the additional feedback (in fact, retention results suggest the additional feedback could be harmful to learning). This refutes the second theory, as the additional feedback did not create more 'memory traces'.

The results suggest a possible modification for either theory. For the first, as the sensorimotor interaction was important for learning when paired with object manipulation, but was not positively impacted by additional audiovisual feedback, the results could be used to argue that sensorimotor encoding is not solely about bodily movement, but also bodily agency, and benefits from contextually and congruent objects for the bodily action to occur with.

For the second theory, these results could be used to suggest that when it comes to sensorimotor interaction, there may be a sweet-spot between too few modalities (no object interaction) and too many (added audiovisual feedback).

5.1.4 IVR, presence, motivation and embodiment

While this thesis is predominantly concerned with sensorimotor interaction, presence, as the preeminent affordance of IVR, was explored through a variety of lenses across the three experiments. Results from the first experiment evidenced that reported feelings of presence correlated with higher learning gains, supporting a learning and presence relationship that is often discussed in literature. However, experiment one also presented evidence that any relationship between learning gain and presence was not mediated by sensorimotor interaction; suggesting that the "two profound affordances" [141] of IVR were acting independently, at least when it comes to IVR language learning.

An interesting discrepancy between experiments was in presence scores and type of sensorimotor interaction. While in the first experiment, there was no presence distinction between the sensorimotor-enabled and non-sensorimotor-enabled conditions, the second experiment present significant differences in both general and spatial presence ratings between the action and gesture interaction conditions. It is odd that no presence distinction was found between between sensorimotor-enabled interaction and non-sensorimotor, when a distinction was found between action and gesture, two conditions that are arguably more similar than in the first experiment.

One explanation for this could be the enhanced sensitivity of the IPQ presence subscales used in the second experiments, which allowed for a more specific examination into types of presence than the single measure presence score used in the first experiment, which may have obfuscated the findings. Another explanation for this discrepancy could be the sample size, with the larger sample of the second experiment able to detect a result that was present but not evidenced in the first experiment. A third explanation could be around the design of the IVR environments: the first experiment used a highly situated cafe setting (which was removed due to some complaints of it being distracting), while the second was set in an abstract location. The highly situated presentation of the first experiment could have led to generally high levels of presence for both conditions, making it harder to detect a distinction. For the second experiment, a participant's feeling of presence could have started at a lower base level due to the abstract environment, and only those who had the benefit of environment interaction (through object manipulation) then began to experience stronger feelings of presence. Finally, the participants in the second experiment likely had much greater IVR experience and exposure than the those in the first experiment, so overall presences levels may have been more sensitive to design choices in the environment, rather than being overriden by high presence from being in a novel IVR experience.

None of these limitations apply to comparisons between the second and third experiments, both of which took place in the abstract setting, and both of which had a large sample size, and both of which used the IPQ. No significant distinction in presence ratings between conditions was found in the third experiment, which suggests that there is a meaningful presence-related distinction between action and gesture, but not high feedback and low feedback conditions. Again, this points to an importance of object manipulation, and not object feedback, reflecting the results regarding learning gain.

The investigations into motivation were less interesting. Despite being highly linked to learning, the results did not show a relationship between learning and motivation, nor any distinctions in motivation between any conditions in any experiment. The presence enquiries varied between experiments, but both the general IMI and learning-game specific MEEGA+ did not show a distinction. While it may be possible to that there was no distinction, it seems more likely that these experimental settings were simply unsuited to finding distinctions in motivation. IVR, experiments in IVR, and language learning in IVR are still relatively novel experiences, and it feels intuitive that these experiments causes universally high feelings of motivation on this novelty.

The results from the third experiment can be used to evidence this view, as the motivation scale was explored as its composite parts: value/usefulness and interest/enjoyment. The interest/enjoyment scale presented no difference between conditions, which evidences the theory that motivation generally was high due to the novelty or IVR medium of the experience, rather than any reflection of the interaction methods.

There was a distinction in the perceived value/usefulness between the third experiments conditions, with the high feedback condition being perceived as more valuable and useful. This perception of value or usefulness was not reflected in learning gain data, which introduces an interesting question regarding whether participants' ratings of value or usefulness are reflective of target outcomes; or whether value/usefulness incorporates aspects other than just learning outcome - and what those aspects are.

Our results also showed that an existing and popular IVR embodied survey was not suitable for reflecting distinctions in sensorimotor interaction conditions. Results from the Gonzalez-Franco embodiment questionnaire [103] in the second experiment did not reflect the sensorimotor interaction conditions. This is a reasonable result, as the survey is focused on understanding embodiment rather than a subsection of it (sensorimotor interaction). In the third experiment, asking participants to rate the interactivity of the experience reflected the distinctions in high and low interactive feedback conditions, while asking about realism or impact on the virtual environment did not. This suggests that future questions investigating sensorimotor interaction could focus around wording such as 'interaction' or 'interactivity'. How this can be deployed to monitor sensorimotor activity, however, remains to be seen.

5.2 Contributions

The research presented in this thesis contributes to multiple research areas: IVR interaction and cognition, theories of embodied cognition; applied linguistics; and to the emerging area of remote IVR experimentation.

Firstly, and primarily, it contributes to the field of IVR interaction and cognition research, with its specific focus on the cognitive impact of HBSI in

IVR on users. For this, it offers four contributions. Firstly, it offers evidence that the decision to leverage HBSI or not can cause different cognitive outcomes (Chapter 5.1.1). Second, it offers evidence that it is not simply the bodily activation that matters, but the interaction with virtual objects to facilitate an outcome, that contributes to the distinct cognitive outcomes (Chapter 4.2). Thirdly, it offers evidence that audiovisual interactional feedback from the IVR system in response to object-based interactions does not provide beneficial cognitive outcomes (in terms of learning), and presents some evidence that this feedback could even be harmful to long-term learning goals (Chapter 4.3). Together, these three contributions form the basis of an argument that foregrounds HBSI as an important factor in the cognitive outcomes from IVR experiences. More specifically, it presents an argument that allowing users to engage in object-based interaction with their hands is an important contributor to IVR experience, and that this interaction and the resulting hand-based agency, is a key interactive aspect of IVR, even if the interacted objects do not provide any additional audiovisual feedback from the interaction.

This research offers a fourth and final contribution to IVR interaction and cognition research, in the form of its scoping review of experimental studies of sensorimotor-based learning in IVR (Chapter 3.1). As well as scoping experimental research on sensorimotor interaction and IVR for what may be the first time, it notes that the field is currently quite disparate and disconnected, despite large overlaps in approach and research justification. It is hoped this contribution will contribute to formalising this research area.

Secondly, this thesis presents evidence supporting embodied theories of cognition, particularly those based upon interactional frameworks. Evidence from the first experiment (Chapter 5.1.1) demonstrate that leveraging bodily activity provides learning benefits over not leveraging it; the second experiment (Chapter 4.2) demonstrates that when the body interacts with an environment (even a virtual one), those outcomes are more potent then body activation without the object-based interaction; and the third experiment (Chapter 4.3) provides evidence that this outcome is not due to additional modes of sensorimotor input and outcome, as additional interactional feedback did not help (and potentially harmed) the learning outcome. Together, these present evidence that cognition is embodied, body-based and interaction-based.

Thirdly, it reinforces some evidence already found in the field of applied linguistics, in which learning verbs with body-based activities out-performs learning without them. It also provides evidence for the distinctions between gesture-based encoding and action-based encoding (Chapter 4.2).

Finally, this thesis provides an overview of the state-of-the-art of conducting IVR experimental research remotely, providing an expert survey of the benefits, drawbacks and remaining questions regarding conducting remote IVR experimentation (Chapter 3.2). This work has already started to form the basis for further work in how to conduct and develop remote XR studies.

5.3 Future work

In this section, six areas for future work are briefly described. Each of these are quite distinct, suggesting larger linguistic investigations beyond verb acquisition (5.3.1); alternative metrics to learning gain (5.3.2); alternative, longitudinal study designs (5.3.3); comparative studies with the physical world (5.3.4); studies into the beyond-real possibilities for sensorimotor interaction in IVR (5.3.5) and investigations into the impact of hand-tracking versus hardware controllers for sensorimotor interaction (5.3.6). Each of these presents avenues to build upon the contributions reported in this thesis.

5.3.1 Learning gain beyond verbs

The results presented in the second experiment show that there can be distinctions in cognitive outcome between sensorimotor interactions that manipulate virtual objects (actions) and sensorimotor interactions that do not manipulate them (gestures).

In this case, that was shown through verb learning, and resulted in improved learning gain for the action interaction type. This limited linguistic scope leaves an important further question: do these sensorimotor distinctions continue to apply to the less-embodied aspects of language learning, such as noun, adjectives and adverbs? There are some hints about this in the first experiment, in which noun learning was not significantly different between a sensorimotor-enabled interaction condition and a non-sensorimotor-enabled condition. However, that experimental sample was quite small, and the experiment was poorly arranged for dedicated noun learning, with nouns being presented after verbs. Therefore whether sensorimotor interaction, and whether the type of interaction (gesture or action) could aid noun, adjective or adverb learning requires further investigation.

Similarly, there are many other academic topics whose learning outcomes could benefit from sensorimotor interaction, but this would have previously been infeasible without the aid of an IVR system. This is starting to be explored [200] but far more work is needed to begin to build a comprehensive picture across many fields.

5.3.2 Factors beyond learning gain

The results in experiment one and two showed significant differences in learning gain between sensorimotor and non-sensorimotor; and sensorimotor (action) and sensorimotor (gesture). However, it failed to detect any other cognitive differences between the conditions despite monitoring presence, motivation and cognitive load.

This suggests either that the type of sensorimotor interaction has no impact on presence, motivation and cognitive load; or that the measures deployed here were not sufficient to detect them. There are certainly more robust ways to examine these factors, both through more objective methods (such as sensors-on-body) as well as through experiment designs better-aimed to investigate these factors. For example, a longitudinal investigation into motivation, with multiple sessions, would be far more robust and accurate than the post-experience motivation questionnaire used in the experiments in this thesis, which only recorded generally high levels of motivation.

There are also other cognitive factors that may be impacted by the sensorimotor interaction type, such as emotional factors (Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment Tze Wei Liew, Nor Azan Mat Zin) which this thesis has not explored. Both identifiying what these factors may be, and then testing them, would be interestingly next steps for this field of study.

5.3.3 Longitudinal investigations and retention

As mentioned above, the study of motivation would benefit from longitudinal investigations. Similarly, language learning retention, another notable outstanding question from this thesis, would also benefit from experiments that ran for a longer period of time. None of the experiments in this thesis found a significant difference in learning retention one week after encoding, for any conditions. Intuitively, that IVR sensorimotor encoding might only provide short-term learning gains seems an unlikely outcome, given the effect size of those gains and the potential reasons (given earlier) for why a significant retention result was not found.

That said, to truly be able to argue that IVR sensorimotor interaction is a useful approach for verb learning, long-term retention evidence needs to be discovered. Typically, for learning to be effectively retained, regular spaced intervals of learning are typically required. Therefore a longitudinal study in which learning is re-examined, re-experienced and re-encoded at regular intervals would provide a rich insight into whether sensorimotor-enabled IVR interactions were a useful approach.

5.3.4 IVR vs Physical World

This thesis based much of its theoretical background on existing research into sensorimotor activity and verb learning gain in the physical world. In the second experiment, the results show a similar relationship between action-based encoding and stronger learning gain outcomes as was found in a physical world study of a similar nature. However, there are few (if any) comparative studies deliberately designed to compare sensorimotor experiments in physical and virtual settings and examine any potential cognitive or experiential differences between the two.

It would be interesting to understand what, if any, distinctions could consistently exist between virtual and physical world explorations. It is clear from research that users do not wholly treat virtual and physical words the same, so the distinctions between the mediums and the implications for IVR sensorimotor experience and outcome would be an interesting area for further study.

5.3.5 IVR: beyond recreations of the physical world

The second experiment presents evidence that the design of sensorimotor interaction in IVR can have impacts on learning outcomes, but through a comparison that has a possible equivalent in the physical world: gestures versus actions. Experiment three, however, begins to examine the design of experience changes that can only occur in IVR environments, comparing high- and low-action feedback. While no distinction was found in learning outcome between these specific conditions, there are a myriad of different aspects of virtual environments that could be customised and changed to potentially impact cognition and learning.

What these could be, what impact they have, and whether we can use these to further push learning beyond what is possible or widely deployed in the physical world, would be an exciting area for further study.

5.3.6 Controllers or hand tracking

The experiments here used physical controllers as inputs - predominantly the Oculus Touch controller. There is evidence, however, that different controllers or controller-free hand-tracking have different effects on user interactions. It is currently unclear if our findings, both language learning and the extrapolations around IVR gestures/actions, are the same when a user is not holding a controller nor pushing a ‘grab’ button, but instead having their hands freely tracked in space. It would be interesting to see a reproduction of any of these studies that explores hand-tracking technology, or comparatively examines the controller and hands-free approaches.

5.4 Closing remarks

The research of this thesis began its life as an investigation into all forms of IVR embodiment and the impact they might have on second language learning. However, it soon became clear that there was little available evidence as to whether the HBSI that users make in IVR, in which they manipulate physical controllers and virtual objects, should be contextualised via theories of gesture, or theories of action, and whether that distinction remained meaningful while interacting in IVR.

The research in this paper supports the view that HBSI provides notable differences compared with non-HBSI, and that those differences are rich enough that acting on virtual objects provides a different cognitive outcome than simply manipulating one's body without manipulating a virtual object. In IVR, HBSI makes a difference, and inside of HBSI, a gesture in IVR is still a gesture, but crucially, an action is not limited to being a gesture.

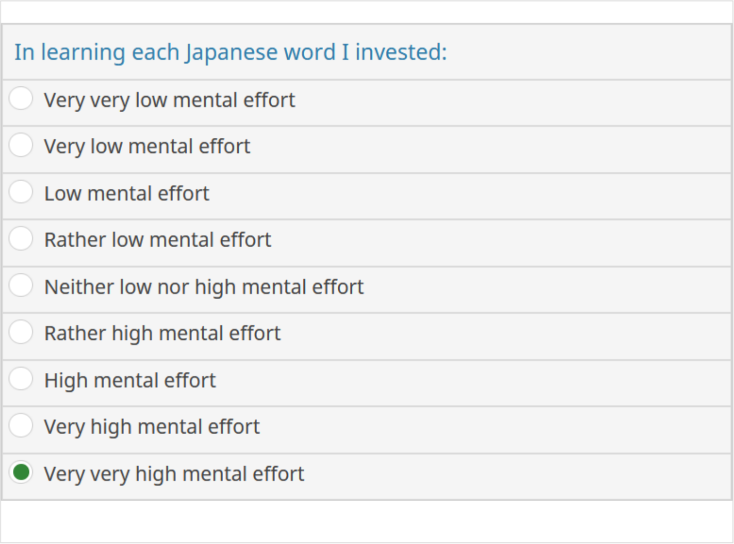
This finding, coupled with the observation in the scoping review that research into sensorimotor interaction and learning in IVR is so disconnected, stresses the need for further research into the impacts of HBSI on users, and for deeper connections between different areas of research (education, human-computer interaction and embodied cognition) examining HBSI in IVR. It is my hope that this work may be received as an early building block for establishing HBSI in IVR as a notable contributor of IVR experiences, that could grow to be as well-explored as head-based sensorimotor interaction and presence.

Additionally, the research on remote IVR user studies may also provide an early resource in this emerging area, with many papers already referencing the published version of this research in their discussions of conducting and reviewing remote IVR studies. In the future, it will be exciting to see what answers will be presented to the outstanding questions for remote IVR experiment procedures outlined in this research.

Finally, I hope that the core findings of this research can lead to further explorations of the natural follow-on questions: what other impacts can HBSI actions have in IVR? What IVR experiences are particularly well-suited to leveraging the cognitive improvements of HBSI? And, most excitingly, what similarities, limitations and benefits can IVR HBSI offer users over existing physical world activities? Because, truly, it is the last of those questions that could have groundbreaking implications for education and society in the centuries to come.

Appendix A

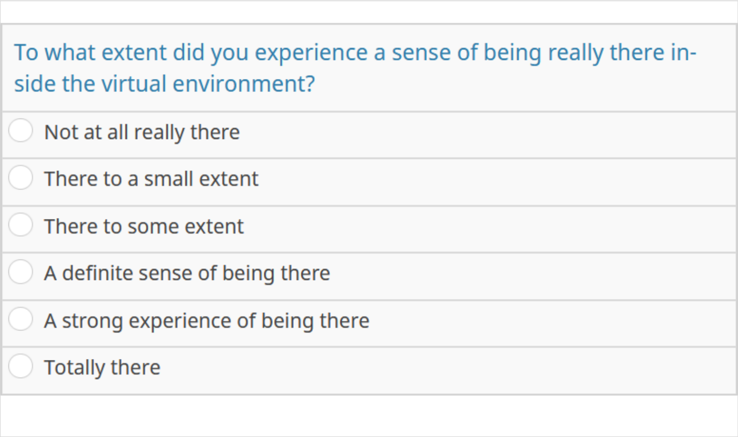
Questionnaires from Experiment #1



In learning each Japanese word I invested:

○ Very very low mental effort
○ Very low mental effort
○ Low mental effort
○ Rather low mental effort
○ Neither low nor high mental effort
○ Rather high mental effort
○ High mental effort
○ Very high mental effort
● Very very high mental effort

Figure A.1: Showing the single-item cognitive load questionnaire, as informed by [231]



To what extent did you experience a sense of being really there inside the virtual environment?

○ Not at all really there

○ There to a small extent

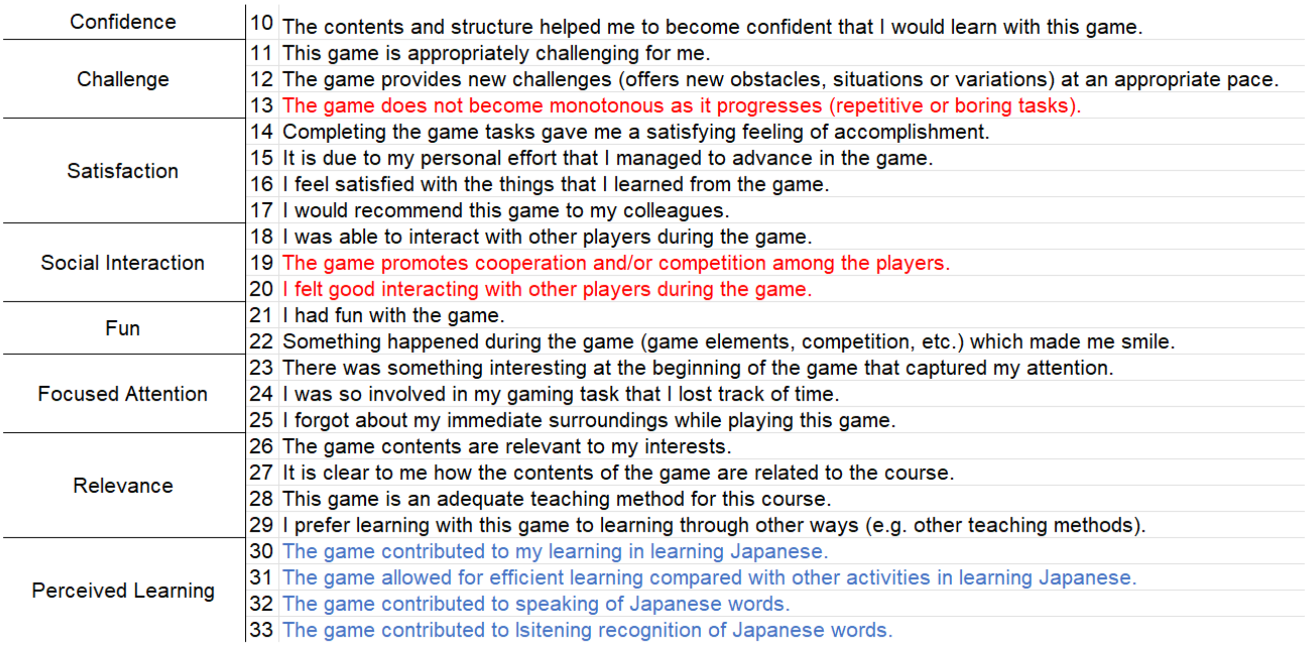
○ There to some extent

○ A definite sense of being there

○ A strong experience of being there

○ Totally there

Figure A.2: Showing the single-item presence questionnaire, as informed by [299]

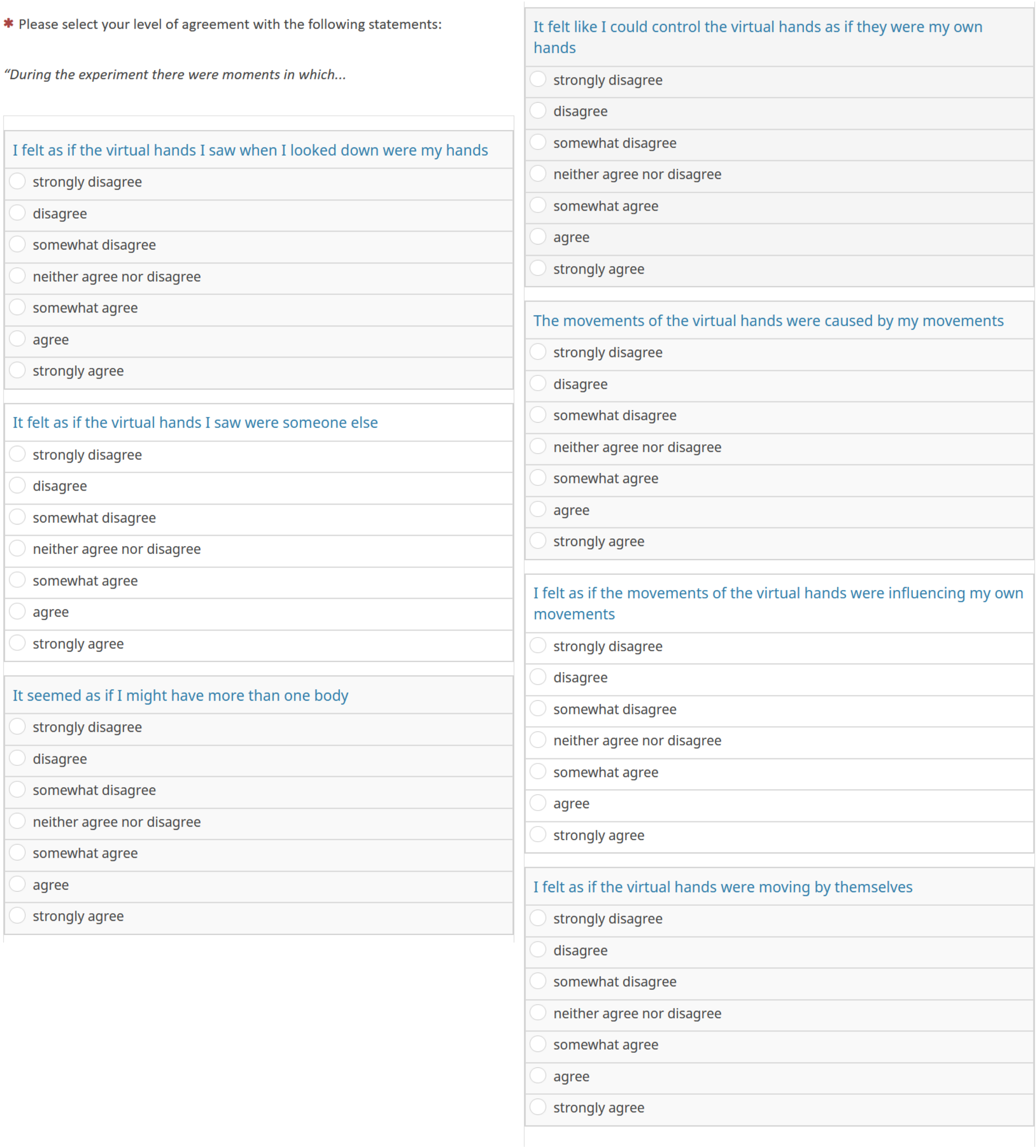


Confidence	10	The contents and structure helped me to become confident that I would learn with this game.
Challenge	11	This game is appropriately challenging for me.
	12	The game provides new challenges (offers new obstacles, situations or variations) at an appropriate pace.
	13	The game does not become monotonous as it progresses (repetitive or boring tasks).
Satisfaction	14	Completing the game tasks gave me a satisfying feeling of accomplishment.
	15	It is due to my personal effort that I managed to advance in the game.
	16	I feel satisfied with the things that I learned from the game.
	17	I would recommend this game to my colleagues.
Social Interaction	18	I was able to interact with other players during the game.
	19	The game promotes cooperation and/or competition among the players.
	20	I felt good interacting with other players during the game.
Fun	21	I had fun with the game.
	22	Something happened during the game (game elements, competition, etc.) which made me smile.
Focused Attention	23	There was something interesting at the beginning of the game that captured my attention.
	24	I was so involved in my gaming task that I lost track of time.
	25	I forgot about my immediate surroundings while playing this game.
Relevance	26	The game contents are relevant to my interests.
	27	It is clear to me how the contents of the game are related to the course.
	28	This game is an adequate teaching method for this course.
	29	I prefer learning with this game to learning through other ways (e.g. other teaching methods).
Perceived Learning	30	The game contributed to my learning in learning Japanese.
	31	The game allowed for efficient learning compared with other activities in learning Japanese.
	32	The game contributed to speaking of Japanese words.
	33	The game contributed to lsitening recognition of Japanese words.

Figure A.3: Showing the questions and categories for the MEEGA+ questionnaire [243]. Questions in red were removed from the survey, while questions in blue were modified to make relevant to the specific experience, as per survey instructions

Appendix B

Questionnaires from Experiment #2



* Please select your level of agreement with the following statements:

"During the experiment there were moments in which...

I felt as if the virtual hands I saw when I looked down were my hands

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

It felt as if the virtual hands I saw were someone else

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

It seemed as if I might have more than one body

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

It felt like I could control the virtual hands as if they were my own hands

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

The movements of the virtual hands were caused by my movements

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

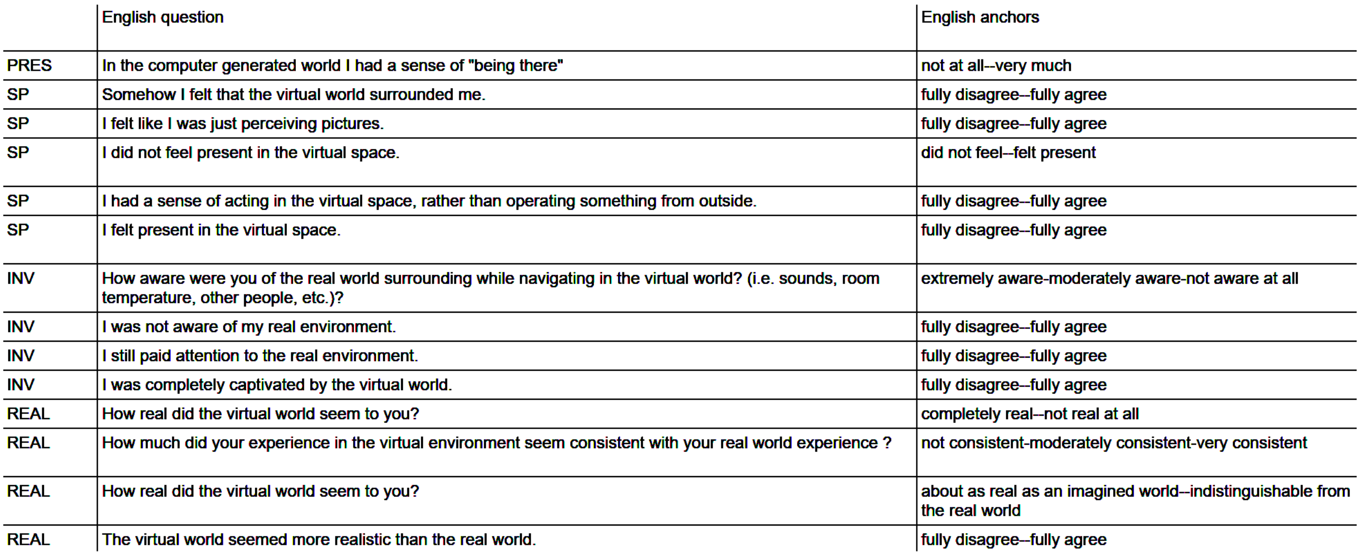
I felt as if the movements of the virtual hands were influencing my own movements

- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

I felt as if the virtual hands were moving by themselves

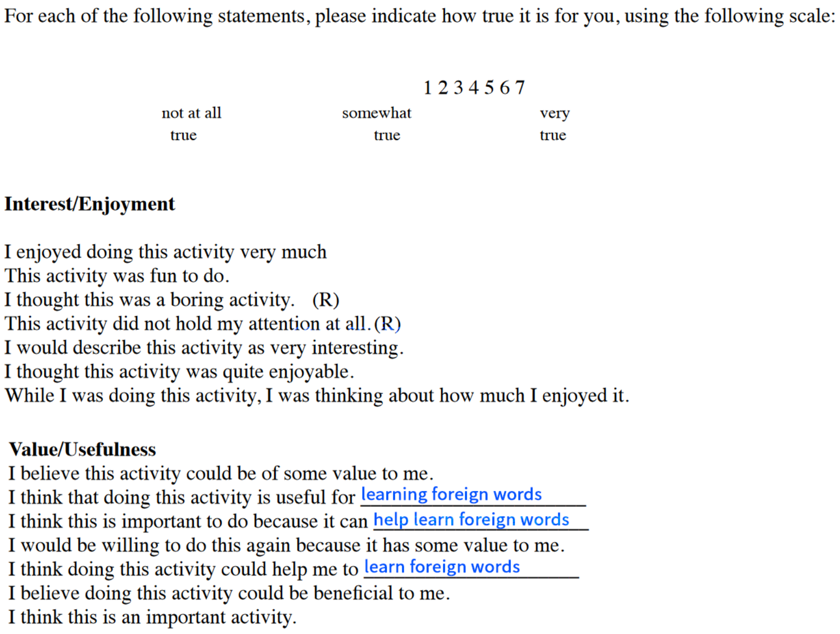
- ○ strongly disagree
- ○ disagree
- ○ somewhat disagree
- ○ neither agree nor disagree
- ○ somewhat agree
- ○ agree
- ○ strongly agree

Figure B.1: Showing the embodiment questionnaire with the embodied agency and body ownership questions, from Gonzalez-Franco [103]



	English question	English anchors
PRES	In the computer generated world I had a sense of "being there"	not at all--very much
SP	Somehow I felt that the virtual world surrounded me.	fully disagree--fully agree
SP	I felt like I was just perceiving pictures.	fully disagree--fully agree
SP	I did not feel present in the virtual space.	did not feel--felt present
SP	I had a sense of acting in the virtual space, rather than operating something from outside.	fully disagree--fully agree
SP	I felt present in the virtual space.	fully disagree--fully agree
INV	How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?	extremely aware-moderately aware-not aware at all
INV	I was not aware of my real environment.	fully disagree--fully agree
INV	I still paid attention to the real environment.	fully disagree--fully agree
INV	I was completely captivated by the virtual world.	fully disagree--fully agree
REAL	How real did the virtual world seem to you?	completely real--not real at all
REAL	How much did your experience in the virtual environment seem consistent with your real world experience ?	not consistent-moderately consistent-very consistent
REAL	How real did the virtual world seem to you?	about as real as an imagined world--indistinguishable from the real world
REAL	The virtual world seemed more realistic than the real world.	fully disagree--fully agree

Figure B.2: Showing the presence types, questions and scale anchors for the Igroup Presence questionnaire [282]



For each of the following statements, please indicate how true it is for you, using the following scale:

1 2 3 4 5 6 7
not at all true
somewhat true
very true

Interest/Enjoyment

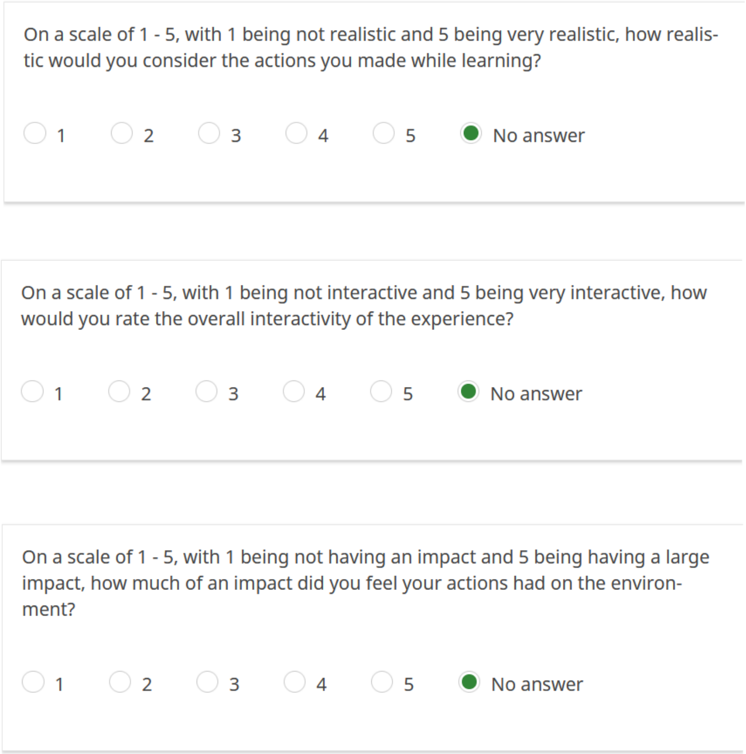
I enjoyed doing this activity very much
This activity was fun to do.
I thought this was a boring activity. (R)
This activity did not hold my attention at all. (R)
I would describe this activity as very interesting.
I thought this activity was quite enjoyable.
While I was doing this activity, I was thinking about how much I enjoyed it.

Value/Usefulness
I believe this activity could be of some value to me.
I think that doing this activity is useful for learning foreign words
I think this is important to do because it can help learn foreign words
I would be willing to do this again because it has some value to me.
I think doing this activity could help me to learn foreign words
I believe doing this activity could be beneficial to me.
I think this is an important activity.

Figure B.3: Showing the Intrinsic Motivation Inventory (IMI) questions for measuring motivation, via the interest/enjoyment and value/usefulness questions [212]. Questions in blue were modified to make relevant to the specific experience, as per survey instructions

Appendix C

Additional questionnaires from Experiment #3



On a scale of 1 - 5, with 1 being not realistic and 5 being very realistic, how realistic would you consider the actions you made while learning?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ● No answer

On a scale of 1 - 5, with 1 being not interactive and 5 being very interactive, how would you rate the overall interactivity of the experience?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ● No answer

On a scale of 1 - 5, with 1 being not having an impact and 5 being having a large impact, how much of an impact did you feel your actions had on the environment?

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ● No answer

Figure C.1: Showing the bespoke questions created to understand if differences in feedback from hand-based sensorimotor interactions in IVR could be detected via questions

Bibliography

[1] D. Abrahamson and R. Lindgren. Embodiment and embodied design. 2014.

[2] L. Q. Allen. The effects of emblematic gestures on the development and access of mental representations of french expressions. *The Modern Language Journal*, 79(4):521–529, 1995.

[3] J. Amores, R. Richer, N. Zhao, P. Maes, and B. M. Eskofier. Promoting relaxation using virtual reality, olfactory interfaces and wearable eeg. In *2018 IEEE 15th international conference on wearable and implantable body sensor networks (BSN)*, pages 98–101. IEEE, 2018.

[4] M. L. Anderson, M. J. Richardson, and A. Chemero. Eroding the boundaries of cognition: Implications of embodiment 1. *Topics in cognitive science*, 4(4):717–730, 2012.

[5] E. Aronson. The theory of cognitive dissonance: A current perspective. *Advances in experimental social psychology*, 4(1):34, 1969.

[6] J. J. Asher. The learning strategy of the total physical response: A review. *The modern language journal*, 50(2):79–84, 1966.

[7] J. J. Asher. The total physical response approach to second language learning. *The modern language journal*, 53(1):3–17, 1969.

[8] F. Aust, B. Diedenhofen, S. Ullrich, and J. Musch. Seriousness checks are useful to improve data validity in online research. *Behavior research methods*, 45(2):527–535, 2013.

[9] L. Bäckman and L.-G. Nilsson. Prerequisites for lack of age differences in memory performance. *Experimental Aging Research*, 11(2):67–73, 1985.

[10] A. D. Baddeley and G. Hitch. Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier, 1974.

[11] A. Badets and M. Pesenti. Creating number semantics through finger movement perception. *Cognition*, 115(1):46–53, 2010.

[12] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S.-H. Jung, and G. Kurillo. The effect of interactivity on learning physical actions in virtual reality. *Media Psychology*, 11(3):354–376, 2008.

[13] J. Bailey, J. N. Bailenson, A. S. Won, J. Flora, and K. C. Armel. Presence and memory: immersive virtual reality effects on cued recall. In *Proceedings of the International Society for Presence Research Annual Conference*, pages 24–26. Citeseer, 2012.

[14] D. Banakou, P. D. Hanumanthu, and M. Slater. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 10:601, 2016.

[15] W. Barfield and C. Hendrix. The effect of update rate on the sense of presence within virtual environments. *Virtual Reality*, 1(1):3–15, 1995.

[16] R. M. Baron and D. A. Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

[17] L. W. Barsalou. Perceptual symbol systems. *Behavioral and brain sciences*, 22(4):577–660, 1999.

[18] O. Baus, S. Bouchard, V. Gougeon, and F.-X. Roucaut. Comparison of anxiety in response to virtual spiders while immersed in augmented reality, head-mounted display, or cave-like system. *Journal of Cybertherapy and Rehabilitation*, 4(2):171–175, 2011.

[19] R. D. Beer. The dynamics of active categorical perception in an evolved model agent. *Adaptive behavior*, 11(4):209–243, 2003.

[20] J. Bennington. Physical controllers vs. hand-and-gesture tracking: Evaluation of control schemes for vr audio mixing. 2019.

[21] K. Bergmann and M. Macedonia. A virtual agent as vocabulary trainer: iconic gestures help to improve learners' memory performance. In *International workshop on intelligent virtual agents*, pages 139–148. Springer, 2013.

[22] M. I. Berkman and E. Akan. Presence and immersion in virtual reality., 2019.

[23] Y. Bian, Q. Wang, C. Zhou, G. Qi, C. Yang, X. Meng, and C. Shen. Comparison of 3d display technologies for embodied interaction in virtual hands-on experiential learning. International Society of the Learning Sciences, Inc.[ISLS]., 2018.

[24] F. Biocca and B. Delaney. Immersive virtual reality technology. *Communication in the age of virtual reality*, 15:32, 1995.

[25] B. Birckhead, C. Khalil, X. Liu, S. Conovitz, A. Rizzo, I. Danovitch, K. Bullock, and B. Spiegel. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. *JMIR mental health*, 6(1):e11973, 2019.

[26] J. Blascovich. Social influence within immersive virtual environments. In *The social life of avatars*, pages 127–145. Springer, 2002.

[27] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological inquiry*, 13(2):103–124, 2002.

[28] J. Boaler, L. Chen, C. Williams, and M. Cordero. Seeing as understanding: The importance of visual mathematics for our brain and learning. *Journal of Applied & Computational Mathematics*, 5(5):1–6, 2016.

[29] A. Borrego, J. Latorre, M. Alcaniz, and R. Llorens. Comparison of oculus rift and htc vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games for health journal*, 7(3):151–156, 2018.

[30] D. A. Bowman, A. Datey, U. Farooq, Y. Ryu, and O. Vasnaik. Empirical comparisons of virtual environment displays. Technical report, Department of Computer Science, Virginia Polytechnic Institute & State ..., 2001.

[31] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[32] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4):571–583, 2007.

[33] D. Brickler, M. Volonte, J. W. Bertrand, A. T. Duchowski, and S. V. Babu. Effects of stereoscopic viewing and haptic feedback, sensory-motor

congruence and calibration on near-field fine motor perception-action coordination in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 28–37. IEEE, 2019.

[34] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq. Studying human behavior with virtual reality: The unity experiment framework. *Behavior research methods*, pages 1–9, 2019.

[35] R. A. Brooks. How to build complete creatures rather than isolated cognitive simulators. In *Architectures for intelligence: the twenty-second Carnegie Mellon Symposium on Cognition*, pages 225–239, 1991.

[36] G. Buccino, I. Colagè, F. Silipo, and P. D'Ambrosio. The concreteness of abstract language: an ancient issue and a new perspective. *Brain Structure and Function*, 224(4):1385–1401, 2019.

[37] T. Buchanan. Potential of the internet for personality research. In *Psychological experiments on the Internet*, pages 121–140. Elsevier, 2000.

[38] L. E. Buck, M. K. Young, and B. Bodenheimer. A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions. *ACM Transactions on Applied Perception (TAP)*, 15(3):1–15, 2018.

[39] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality data? 2016.

[40] M. D. Buhrmester, S. Talaifar, and S. D. Gosling. An evaluation of amazon's mechanical turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, 13(2):149–154, 2018.

[41] F. Buttussi and L. Chittaro. Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics*, 24(2):1063–1076, 2017.

[42] M. Canale and M. Swain. Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1):1–47, 1980.

[43] K. Casler, L. Bickel, and E. Hackett. Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing. *Computers in human behavior*, 29(6):2156–2160, 2013.

[44] M. Celce-Murcia. Rethinking the role of communicative competence in language teaching. In *Intercultural language use and language learning*, pages 41–57. Springer, 2008.

[45] M. Celce-Murcia, Z. Dörnyei, and S. Thurrell. Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied linguistics*, 6(2):5–35, 1995.

[46] T. Chandrasekera, K. Fernando, and L. Puig. Effect of degrees of freedom on the sense of presence generated by virtual reality (vr) head-mounted display systems: a case study on the use of vr in early design studios. *Journal of Educational Technology Systems*, 47(4):513–522, 2019.

[47] J. S.-K. Chang, G. Yeboah, A. Doucette, P. Clifton, M. Nitsche, T. Welsh, and A. Mazalek. Evaluating the effect of tangible virtual reality on spatial perspective taking ability. In *Proceedings of the 5th symposium on spatial user interaction*, pages 68–77, 2017.

[48] K.-J. Chao, H.-W. Huang, W.-C. Fang, and N.-S. Chen. Embodied play to learn: Exploring k inect-facilitated memory performance. *British Journal of Educational Technology*, 44(5):E151–E155, 2013.

[49] D. Checa and A. Bustillo. A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 79(9):5501–5527, 2020.

[50] A. Cheng, L. Yang, and E. Andersen. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 541–549, 2017.

[51] J. A. Cheyne. The ominous numinous. sensed presence and'other'hallucinations. *Journal of Consciousness Studies*, 8(5-6):133–150, 2001.

[52] A. W. Chickering and G. D. Kuh. Promoting student success: Creating conditions so every student can learn. 2005.

[53] H. J. Chiel and R. D. Beer. The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, 20(12):553–557, 1997.

[54] B. R. Chiswick and P. W. Miller. Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development*, 26(1):1–11, 2005.

[55] L. Chittaro and F. Buttussi. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE transactions on visualization and computer graphics*, 21(4):529–538, 2015.

[56] W. Choi, L. Li, S. Satoh, and K. Hachimura. Multisensory integration in the virtual hand illusion with active movement. *BioMed research international*, 2016, 2016.

[57] M. Christofi, D. Michael-Grigoriou, and C. Kyrlitsias. A virtual reality simulation of drug users' everyday life: the effect of supported sensorimotor contingencies on empathy. *Frontiers in psychology*, 11:1242, 2020.

[58] A. Clark. *Being there: Putting brain, body, and world together again.* MIT press, 1998.

[59] A. Clark et al. *Supersizing the mind: Embodiment, action, and cognitive extension.* OUP USA, 2008.

[60] R. E. Clark. Reconsidering research on learning from media. *Review of educational research*, 53(4):445–459, 1983.

[61] R. E. Clark and T. G. Craig. Research and theory on multi-media learning effects. In *Interactive multimedia learning environments*, pages 19–30. Springer, University of Southern California USA, 1992.

[62] D. H. Clements. 'concrete'manipulatives, concrete ideas. *Contemporary issues in early childhood*, 1(1):45–60, 2000.

[63] R. L. Cohen. Memory for action events: The power of enactment. *Educational psychology review*, 1(1):57–80, 1989.

[64] L. Connelly, Y. Jia, M. L. Toro, M. E. Stoykov, R. V. Kenyon, and D. G. Kamper. A pneumatic glove and immersive virtual reality environment for hand rehabilitative training after stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(5):551–559, 2010.

[65] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE transactions on visualization and computer graphics*, 23(1):441–450, 2016.

[66] J. H. Corpus, M. S. McClintic-Gilbert, and A. O. Hayenga. Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology*, 34(2):154–166, 2009.

[67] S. A. Creasy, R. J. Rogers, K. K. Davis, B. B. Gibbs, E. E. Kershaw, and J. M. Jakicic. Effects of supervised and unsupervised physical activity programmes for weight loss. *Obesity Science & Practice*, 3(2):143–152, 2017.

[68] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142, 1993.

[69] J. J. Cummings and J. N. Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2):272–309, 2016.

[70] T. de Back, R. van Hoef, A. M. Tinga, and M. M. Louwerse. Presence is key: Unlocking performance benefits of immersive virtual reality. In *CogSci*, 2018.

[71] J. A. de Nooijer, T. van Gog, F. Paas, and R. A. Zwaan. Words in action: using gestures to improve verb learning in primary school children. *Gesture*, 14(1):46–69, 2014.

[72] J. Dewey. My pedagogic creed. *Journal of Education*, 101(18):490–490, 1925.

[73] A. F. Di Natale, C. Repetto, G. Riva, and D. Villani. Immersive virtual reality in k-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, 51(6):2006–2033, 2020.

[74] D. Edge, K.-Y. Cheng, and M. Whitney. Spatialease: learning language through body motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 469–472. ACM, 2013.

[75] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray. An evaluation of heart rate and electrodermal activity as an objective qoe evaluation method for immersive virtual reality environments. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2016.

[76] S. B. Ehrlich, S. C. Levine, and S. Goldin-Meadow. The importance of gesture in children's spatial reasoning. *Developmental psychology*, 42(6):1259, 2006.

[77] A. Elor, M. Powell, E. Mahmoodi, N. Hawthorne, M. Teodorescu, and S. Kurniawan. On shooting stars: Comparing cave and hmd immersive virtual reality exergaming for adults with mixed ability. *ACM Transactions on Computing for Healthcare*, 1(4):1–22, 2020.

[78] D. P. Ely and B. B. Minor. *Educational Media and Technology Yearbook, 1994. Volume 20.* ERIC, 1994.

[79] J. Engelkamp and H. Krumnacker. Image-and motor-processes in the retention of verbal materials. *Zeitschrift für experimentelle und angewandte Psychologie*, 1980.

[80] J. Epelboim. Deictic codes, embodiment of cognition, and the real world. *Behavioral and Brain Sciences*, 20(4):746–746, 1997.

[81] S. W. Eskildsen and J. Wagner. Embodied l2 construction learning. *Language Learning*, 65(2):268–297, 2015.

[82] Z. Feng, V. A. González, R. Amor, R. Lovreglio, and G. Cabrera-Guerrero. Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education*, 127:252–266, 2018.

[83] L. Figueiredo, E. Rodrigues, J. Teixeira, and V. Teichrieb. A comparative evaluation of direct hand and wand interactions on consumer devices. *Computers & Graphics*, 77:108–121, 2018.

[84] N. D. Finkelstein, W. K. Adams, C. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. LeMaster. When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical review special topics-physics education research*, 1(1):010103, 2005.

[85] N. Fleming and C. Mills. Vark: A guide to learning styles. 2001.

[86] N. D. Fleming. *Teaching and learning styles: VARK strategies.* IGI global, 2001.

[87] M. C. Frank, J. B. Tenenbaum, and E. Gibson. Learning and long-term retention of large-scale artificial languages. *PloS one*, 8(1):e52500, 2013.

[88] S. Franklin, S. Wolpert, S. R. McKay, and W. Christian. Artificial minds. *Computers in Physics*, 11(3):258–259, 1997.

[89] D. Freeman, P. Haselton, J. Freeman, B. Spanlang, S. Kishore, E. Albery, M. Denne, P. Brown, M. Slater, and A. Nickless. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, 5(8):625–632, 2018.

[90] O. Fuhrman, A. Eckerling, N. Friedmann, R. Tarrasch, and G. Raz. The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning*, 2020.

[91] O. Fuhrman, A. Eckerling, N. Friedmann, R. Tarrasch, and G. Raz. The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning*, 37(3):672–683, 2021.

[92] A. Gaggioli, M. Bassi, and A. Fave. Quality of experience in virtual environments. *Emerging Communication*, 5:121–136, 2003.

[93] M. Garau, D. Friedman, H. R. Widenfeld, A. Antley, A. Brogni, and M. Slater. Temporal and spatial variations in presence: Qualitative analysis of interviews from an experiment on breaks in presence. *Presence: Teleoperators and Virtual Environments*, 17(3):293–309, 2008.

[94] A. Glenberg, J. Willford, B. Gibson, A. Goldberg, and X. Zhu. Improving reading to improve math. *Scientific Studies of Reading*, 16(4):316–340, 2012.

[95] A. M. Glenberg. What memory is for. *Behavioral and brain sciences*, 20(1):1–19, 1997.

[96] A. M. Glenberg, A. B. Goldberg, and X. Zhu. Improving early reading comprehension using embodied cai. *Instructional Science*, 39(1):27–39, 2011.

[97] A. M. Glenberg, T. Gutierrez, J. R. Levin, S. Japuntich, and M. P. Kaschak. Activity and imagined activity can enhance young children's reading comprehension. *Journal of educational psychology*, 96(3):424, 2004.

[98] A. M. Glenberg and M. P. Kaschak. Grounding language in action. *Psychonomic bulletin & review*, 9(3):558–565, 2002.

[99] J. A. Glover. The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81(3):392, 1989.

[100] S. Goldin-Meadow and S. L. Beilock. Action's influence on thought: The case of gesture. *Perspectives on psychological science*, 5(6):664–674, 2010.

[101] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik. Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer assisted language learning*, 27(1):70–105, 2014.

[102] J. González, A. Barros-Loscertales, F. Pulvermüller, V. Meseguer, A. Sanjuán, V. Belloch, and C. Ávila. Reading cinnamon activates olfactory brain regions. *Neuroimage*, 32(2):906–912, 2006.

[103] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[104] M. A. Goodale, A. D. Milner, et al. Separate visual pathways for perception and action. 1992.

[105] J. K. Goodman, C. E. Cryder, and A. Cheema. Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224, 2013.

[106] W. Goodrich and C. L. Hudson Kam. Co-speech gesture as input in verb learning. *Developmental science*, 12(1):81–87, 2009.

[107] C. L. Gordon, T. M. Shea, D. C. Noelle, and R. Balasubramaniam. Affordance compatibility effect for word learning in virtual reality. *Cognitive science*, 43(6):e12742, 2019.

[108] S. D. Gosling, S. Vazire, S. Srivastava, and O. P. John. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American psychologist*, 59(2):93, 2004.

[109] S. T. Grafton, L. Fadiga, M. A. Arbib, and G. Rizzolatti. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*, 6(4):231–236, 1997.

[110] S. Grassini, K. Laumann, and M. R. Skogstad. The use of virtual reality alone does not promote training performance (but sense of presence does). *Frontiers in Psychology*, 11, 2020.

[111] R. Gray. Transfer of training from virtual to real baseball batting. *Frontiers in psychology*, 8:2183, 2017.

[112] D. Gromala, X. Tong, A. Choo, M. Karamnejad, and C. D. Shaw. The virtual meditative walk: virtual reality therapy for chronic pain management. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 521–524, 2015.

[113] A. Grysman. Collecting narrative data on amazon's mechanical turk. *Applied Cognitive Psychology*, 29(4):573–583, 2015.

[114] S. G. Hart. Nasa task load index (tlx). volume 1.0; paper and pencil package. 1986.

[115] J. Hartfill, J. Gabel, D. Neves-Coelho, D. Vogel, F. Räthel, S. Tiede, O. Ariza, and F. Steinicke. Word saber: an effective and fun vr vocabulary learning game. In *Proceedings of the Conference on Mensch und Computer*, pages 145–154, 2020.

[116] O. Hauk, I. Johnsrude, and F. Pulvermüller. Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2):301–307, 2004.

[117] P. Havig, J. McIntire, and E. Geiselman. Virtual reality in a cave: limitations and the need for hmds? In *Head-and Helmet-Mounted Displays XVI: Design and Applications*, volume 8041, page 804107. International Society for Optics and Photonics, 2011.

[118] D. Hecht, M. Reiner, and G. Halevy. Multimodal virtual environments: response times, attention, and presence. *Presence: Teleoperators and virtual environments*, 15(5):515–523, 2006.

[119] J. Henrich, S. J. Heine, and A. Norenzayan. Most people are not weird. *Nature*, 466(7302):29–29, 2010.

[120] M. Hertzum, P. Borlund, and K. B. Kristoffersen. What do thinking-aloud participants say? a comparison of moderated and unmoderated usability sessions. *International Journal of Human-Computer Interaction*, 31:557 – 570, 2015.

[121] S. A. Hillen, K. U. Breuer, and R. D. Tennyson. Gaming and learning: Theory, research, and practice. *Challenges Facing Contemporary Didactics: Diversity of Students and the Role of New Media in Teaching and Learning*, page 185, 2011.

[122] J. H. Ho, S. Z. Zhou, D. Wei, and A. Low. Investigating the effects of educational game with wii remote on outcomes of learning. In *Transactions on Edutainment III*, pages 240–252. Springer, 2009.

[123] M. Hofer, T. Hartmann, A. Eden, R. Ratan, and L. Hahn. The role of plausibility in the experience of spatial presence in virtual environments. *Frontiers in Virtual Reality*, 1:2, 2020.

[124] H. G. Hoffman, A. Garcia-Palacios, V. Kapa, J. Beecher, and S. R. Sharar. Immersive virtual reality for reducing experimental ischemic pain. *International Journal of Human-Computer Interaction*, 15(3):469–486, 2003.

[125] S. L. Hornstein and N. W. Mulligan. Memory of action events: The role of objects in memory of self-and other-performed tasks. *The American journal of psychology*, 114(2):199, 2001.

[126] A. B. Hostetter and M. W. Alibali. Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, 15(3):495–514, 2008.

[127] A. B. Hostetter, W. Pouw, and E. M. Wakefield. Learning from gesture and action: An investigation of memory for where objects went and how they got there. *Cognitive Science*, 44(9):e12889, 2020.

[128] M. C. Howard. Virtual reality interventions for personal development: A meta-analysis of hardware and software. *Human–Computer Interaction*, 34(3):205–239, 2019.

[129] C. L. Hoyt, J. Blascovich, and K. R. Swinth. Social inhibition in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(2):183–195, 2003.

[130] W. Huang, R. D. Roscoe, M. C. Johnson-Glenberg, and S. D. Craig. Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, 37(3):745–758, 2021.

[131] B. Huber and K. Z. Gajos. Conducting online virtual environment experiments with uncompensated, unsupervised samples. *Plos one*, 15(1):e0227629, 2020.

[132] J. H. Hulstijn. Second language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition*, 19(2):131–143, 1997.

[133] J.-P. Huttner, D. Pfeiffer, and S. Robra-Bissantz. Imaginary versus virtual loci: evaluating the memorization accuracy in a virtual memory palace. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[134] D. Hymes. On communicative competence. *sociolinguistics*, 269293:269–293, 1972.

[135] W. A. IJsselsteijn, H. De Ridder, J. Freeman, and S. E. Avons. Presence: concept, determinants, and measurement. In *Human vision and electronic imaging V*, volume 3959, pages 520–529. International Society for Optics and Photonics, 2000.

[136] D. Ionescu, B. Ionescu, C. Gadea, and S. Islam. Multimodal control of virtual game environments through gestures and physical controllers. In *2011 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings*, pages 1–6. IEEE, 2011.

[137] D. Jacobson. On theorizing presence. *Journal of Virtual Environments*, 6(1):1–8, 2002.

[138] L. Jensen and F. Konradsen. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4):1515–1529, 2018.

[139] T. Jia and Y. Liu. Words in kitchen: An instance of leveraging virtual reality technology to learn vocabulary. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 150–155. IEEE, 2019.

[140] D. Jirak, M. M. Menz, G. Buccino, A. M. Borghi, and F. Binkofski. Grasping language–a short story on embodiment. *Consciousness and cognition*, 19(3):711–720, 2010.

[141] M. C. Johnson-Glenberg. Immersive vr and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5:81, 2018.

[142] M. C. Johnson-Glenberg, D. A. Birchfield, L. Tolentino, and T. Koziupa. Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, 106(1):86, 2014.

[143] M. C. Johnson-Glenberg, C. Megowan-Romanowicz, D. A. Birchfield, and C. Savio-Ramos. Effects of embodied learning and digital platform on the retention of physics content: Centripetal force. *Frontiers in psychology*, 7:1819, 2016.

[144] N. Josman, E. Somer, A. Reisberg, P. L. Weiss, A. Garcia-Palacios, and H. Hoffman. Busworld: designing a virtual environment for post-traumatic stress disorder in israel: a protocol. *Cyberpsychology & Behavior*, 9(2):241–244, 2006.

[145] M. C. Juan and D. Pérez. Comparison of the levels of presence and anxiety in an acrophobic environment viewed via hmd or cave. *Presence: Teleoperators and virtual environments*, 18(3):232–248, 2009.

[146] M. Juriševič, S. A. Glažar, C. R. Pučko, and I. Devetak. Intrinsic motivation of pre-service primary school teachers for learning chemistry in relation to their academic achievement. *International Journal of Science Education*, 30(1):87–107, 2008.

[147] T. Kahlert, F. v. d. Camp, and R. Stiefelhagen. Learning to juggle in an interactive virtual reality environment. In *International Conference on Human-Computer Interaction*, pages 196–201. Springer, 2015.

[148] E. Kalina and M. C. Johnson-Glenberg. Presence and platform: Effects of embodiment comparing a 2d computer and 3d vr game. In *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*, pages 31–37. IEEE, 2020.

[149] M. P. Kaschak, C. M. Connor, and J. L. Dombek. Enacted reading comprehension: using bodily movement to aid the comprehension of abstract text content. *PloS one*, 12(1):e0169711, 2017.

[150] F. Ke, S. Lee, and X. Xu. Teaching training in a mixed-reality integrated learning environment. *Computers in Human Behavior*, 62:212–220, 2016.

[151] Y. Ke, H. Qi, F. He, S. Liu, X. Zhao, P. Zhou, L. Zhang, and D. Ming. An eeg-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. *Frontiers in human neuroscience*, 8:703, 2014.

[152] R. Kehl and L. Van Gool. Real-time pointing gesture recognition for an immersive environment. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pages 577–582. IEEE, 2004.

[153] S. Kelly, M. Healey, A. Özyürek, and J. Holler. The processing of speech, gesture, and action during language comprehension. *Psychonomic bulletin & review*, 22(2):517–523, 2015.

[154] S. D. Kelly, S. M. Manning, and S. Rodak. Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, 2(4):569–588, 2008.

[155] S. D. Kelly, T. McDevitt, and M. Esch. Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and cognitive processes*, 24(2):313–334, 2009.

[156] S. D. Kelly, A. Özyürek, and E. Maris. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2):260–267, 2010.

[157] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[158] L. Kerswell, P. D. Siakaluk, P. M. Pexman, C. R. Sears, and W. J. Owen. Homophone effects in visual word recognition depend on homophone type and task demands. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 61(4):322, 2007.

[159] P. Kettunen and J. Oksanen. Effects of unsupervised participation over the internet on a usability study about map animation. 2018.

[160] F. Keusch. Why do people participate in web surveys? applying survey participation theory to internet survey data collection. *Management review quarterly*, 65(3):183–216, 2015.

[161] K. Kilteni, I. Bergstrom, and M. Slater. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE transactions on visualization and computer graphics*, 19(4):597–605, 2013.

[162] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[163] K. Kim, M. Z. Rosenthal, D. Zielinski, and R. Brady. Comparison of desktop, head mounted display, and six wall fully immersive systems using a stressful task. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 143–144. IEEE, 2012.

[164] D. Kirsh and P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive science*, 18(4):513–549, 1994.

[165] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. 2007.

[166] J. Kjeldskov. Interaction: Full and partial immersive virtual reality displays. *IRIS24*, pages 587–600, 2001.

[167] R. Kormi-Nouri and L.-G. Nilsson. The motor component. *Memory for action: A distinct form of episodic memory?*, page 97, 2001.

[168] P. Kourtesis, D. Korre, S. Collina, L. A. Doumas, and S. E. MacPherson. Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of virtual reality everyday assessment lab (vr-eal), a neuropsychological test battery in immersive virtual reality. *Frontiers in Computer Science*, 1:12, 2020.

[169] J. H. Krantz and R. Dalal. Validity of web-based psychological research. In *Psychological experiments on the Internet*, pages 35–60. Elsevier, 2000.

[170] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann. Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality. In *Proceedings of the 12th acm international conference on pervasive technologies related to assistive environments*, pages 289–298, 2019.

[171] E. Krokos, C. Plaisant, and A. Varshney. Virtual memory palaces: immersion aids recall. *Virtual reality*, 23(1):1–15, 2019.

[172] A. Lacroix, R. Kressig, T. Muehlbauer, Y. Gschwind, B. Pfenninger, O. Bruegger, and U. Granacher. Effects of a supervised versus an unsupervised combined balance and strength training program on balance and muscle power in healthy older adults: A randomized controlled trial. *Gerontology*, 62:275 – 288, 2015.

[173] R. Lammfromm and D. Gopher. Transfer of skill from a virtual reality trainer to real juggling. In *BIO web of conferences*, volume 1, page 00054. EDP Sciences, 2011.

[174] R. W. Langacker. *Foundations of cognitive grammar: Theoretical prerequisites*, volume 1. Stanford university press, 1987.

[175] E. A.-L. Lee, K. W. Wong, and C. C. Fung. How does desktop virtual reality enhance learning outcomes? a structural equation modeling approach. *Computers & Education*, 55(4):1424–1442, 2010.

[176] S. Lee, K. Park, J. Lee, and K. Kim. User study of vr basic controller and data glove as hand gesture inputs in vr games. In *2017 international symposium on ubiquitous virtual reality (isuvr)*, pages 1–3. IEEE, 2017.

[177] W.-J. Lee, C.-W. Huang, C.-J. Wu, S.-T. Huang, and G.-D. Chen. The effects of using embodied interactions to improve learning performance. In *2012 IEEE 12th International Conference on Advanced Learning Technologies*, pages 557–559. IEEE, 2012.

[178] M. R. Lepper. Motivational considerations in the study of instruction. *Cognition and instruction*, 5(4):289–309, 1988.

[179] J. R. Lewis. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018.

[180] G. Li, H. Tang, Y. Sun, J. Kong, G. Jiang, D. Jiang, B. Tao, S. Xu, and H. Liu. Hand gesture recognition based on convolution neural network. *Cluster Computing*, 22(2):2719–2729, 2019.

[181] Y. Li, T. Ma, and E. Ch'Ng. Enhancing vr experiential learning through the design of embodied interaction in a shared virtual environment. In *PACIS*, page 185, 2019.

[182] C.-Y. Lin, Y.-H. Jen, L.-C. Wang, H.-H. Lin, and L.-W. Chang. Assessment of the application of wii remote for the design of interactive teaching materials. In *International Conference on Information and Management Engineering*, pages 483–490. Springer, 2011.

[183] R. Lindgren and M. Johnson-Glenberg. Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, 42(8):445–452, 2013.

[184] S. E. Lindley, J. Le Couteur, and N. L. Berthouze. Stirring up experience through movement in game play: effects on engagement and social behaviour. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 511–514. ACM, 2008.

[185] Y. Liu, Z. Liu, and Y. Jin. Virtual dome system using hmds: An alternative to the expensive and less accessible physical domes. In *International Conference on Human-Computer Interaction*, pages 312–328. Springer, 2019.

[186] G. R. Loftus. Comparison of recognition and recall in a continuous memory task. *Journal of experimental psychology*, 91(2):220, 1971.

[187] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2):JCMC321, 1997.

[188] M. Lombard and M. T. Jones. Defining presence. In *Immersed in Media*, pages 13–34. Springer, 2015.

[189] J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, 31(4):557–564, 1999.

[190] B. L. Ludlow. Virtual reality: Emerging applications and future directions. *Rural Special Education Quarterly*, 34(3):3–10, 2015.

[191] X. Ma, M. Cackett, L. Park, E. Chien, and M. Naaman. Web-based vr experiments powered by the crowd. In *Proceedings of the 2018 World Wide Web Conference*, pages 33–43, 2018.

[192] M. Macedonia. Embodied learning: Why at school the mind needs the body. *Frontiers in Psychology*, 10:2098, 2019.

[193] M. Macedonia and T. R. Knösche. Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, 5(4):196–211, 2011.

[194] M. Macedonia, A. Lehner, and C. Repetto. Positive effects of grasping virtual objects on memory for novel words in a second language. *Scientific Reports*, 10(1):1–13, 2020.

[195] M. Macedonia and K. Mueller. Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in psychology*, 7:953, 2016.

[196] M. Macedonia, C. Repetto, A. Ischebeck, and K. Mueller. Depth of encoding through observed gestures in foreign language word learning. *Frontiers in psychology*, 10, 2019.

[197] C. M. MacLeod, N. Gopie, K. L. Hourihan, K. R. Neary, and J. D. Ozubko. The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3):671, 2010.

[198] S. L. Macrine and J. M. Fugate. Embodied cognition. In *Oxford Research Encyclopedia of Education*. Oxford Research Encyclopedias, 2020.

[199] C. R. Madan and A. Singhal. Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition*, 21(3):1563–1570, 2012.

[200] J. H. Madden, A. S. Won, J. P. Schuldt, B. Kim, S. Pandita, Y. Sun, T. Stone, and N. Holmes. Virtual reality as a teaching tool for moon phases and beyond. *arXiv preprint arXiv:1807.11179*, 2018.

[201] G. Makransky, S. Borre-Gude, and R. E. Mayer. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 2019.

[202] G. Makransky, L. Lilleholt, and A. Aaby. Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, 72:276–285, 2017.

[203] G. Makransky and G. B. Petersen. The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3):937–958, 2021.

[204] G. Makransky, T. S. Terkildsen, and R. E. Mayer. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60:225–236, 2019.

[205] J. A. Mangels and A. Heinberg. Improved episodic integration through enactment: Implications for aging. *The Journal of General Psychology*, 133(1):37–65, 2006.

[206] C. Maraj, J. Hurter, S. Ferrante, L. Horde, J. Carter, and S. Murphy. Oculus rift versus htc vive: Usability assessment from a teleportation task. In *International Conference on Human-Computer Interaction*, pages 247–257. Springer, 2019.

[207] D. M. Markowitz, R. Laha, B. P. Perone, R. D. Pea, and J. N. Bailenson. Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in Psychology*, 9:2364, 2018.

[208] T. Martin and D. L. Schwartz. Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive science*, 29(4):587–625, 2005.

[209] W. Mason and S. Suri. Conducting behavioral research on amazon's mechanical turk. *Behavior research methods*, 44(1):1–23, 2012.

[210] K. Masumoto, M. Yamaguchi, K. Sutani, S. Tsuneto, A. Fujita, and M. Tonoike. Reactivation of physical motor information in the memory of action events. *Brain research*, 1101(1):102–109, 2006.

[211] R. E. Mayer. Multimedia learning. In *Psychology of learning and motivation*, volume 41, pages 85–139. Elsevier, 2002.

[212] E. McAuley, T. Duncan, and V. V. Tammen. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1):48–58, 1989.

[213] S. G. McCafferty. Space for cognition: Gesture and second language learning. *International Journal of Applied Linguistics*, 14(1):148–165, 2004.

[214] D. McNeill. Gesture, gaze, and ground. In *International workshop on machine learning for multimodal interaction*, pages 1–14. Springer, 2005.

[215] K. Meyerbröker, N. Morina, G. Kerkhof, P. M. Emmelkamp, et al. Virtual reality exposure treatment of agoraphobia: A comparison of computer automatic virtual environment and head-mounted display. *Annual Review of Cybertherapy and Telemedicine*, 9(1):41–45, 2011.

[216] T. A. Mikropoulos. Presence: a unique characteristic in educational virtual environments. *Virtual Reality*, 10(3-4):197–206, 2006.

[217] T. A. Mikropoulos and A. Natsis. Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3):769–780, 2011.

[218] R. Mitchell, F. Myles, and E. Marsden. *Second language learning theories*. Routledge, 2019.

[219] Y. Moon. The effects of distance in local versus remote human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 103–108, 1998.

[220] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012.

[221] J. D. Moss and E. R. Muth. Characteristics of head-mounted displays and their effects on simulator sickness. *Human factors*, 53(3):308–319, 2011.

[222] A. Mottelson and K. Hornbæk. Virtual reality studies outside the laboratory. In *Proceedings of the 23rd acm symposium on virtual reality software and technology*, pages 1–10, 2017.

[223] A. Müller-Hartmann and M. Schocker-von Ditfurth. *Teaching English: Task-supported language learning*, volume 3336. UTB, 2011.

[224] K. H. Mumford and S. Kita. Children use gesture to interpret novel verb meanings. *Child Development*, 85(3):1181–1189, 2014.

[225] L. Murtagh and F. van der Slik. Retention of irish skills: A longitudinal study of a school-acquired second language. *International Journal of Bilingualism*, 8(3):279–302, 2004.

[226] M. A. Novack, E. M. Wakefield, and S. Goldin-Meadow. What makes a movement a gesture? *Cognition*, 146:339–348, 2016.

[227] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

[228] R. E. Núñez, L. D. Edwards, and J. F. Matos. Embodied cognition as grounding for situatedness and context in mathematics education. *Educational studies in mathematics*, 39(1-3):45–65, 1999.

[229] L. Nyberg. *The enactment effect: Studies of a memory phenomenon*. PhD thesis, Umeå Universitet, 1993.

[230] Ş. Özçalışkan and S. Goldin-Meadow. Gesture is at the cutting edge of early language development. *Cognition*, 96(3):B101–B113, 2005.

[231] F. G. Paas. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, 84(4):429, 1992.

[232] F. G. Paas and J. J. Van Merriënboer. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology*, 86(1):122, 1994.

[233] X. Pan and A. F. d. C. Hamilton. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3):395–417, 2018.

[234] G. Paolacci and J. Chandler. Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188, 2014.

[235] G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[236] M. Papastergiou. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & education*, 52(1):1–12, 2009.

[237] D. Parmar, J. Isaac, S. V. Babu, N. D'Souza, A. E. Leonard, S. Jörg, K. Gundersen, and S. B. Daily. Programming moves: Design and evaluation of applying embodied interaction in virtual environments to enhance computational thinking in middle school students. In *2016 IEEE Virtual Reality (VR)*, pages 131–140. IEEE, 2016.

[238] J. Parong and R. E. Mayer. Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6):785, 2018.

[239] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

[240] E. Peer, J. Vosgerau, and A. Acquisti. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior research methods*, 46(4):1023–1031, 2014.

[241] S. Persky, K. A. Kaphingst, C. McCall, C. Lachance, A. C. Beall, and J. Blascovich. Presence relates to distinct outcomes in two virtual environments employing different learning modalities. *CyberPsychology & Behavior*, 12(3):263–268, 2009.

[242] G. B. Petersen, G. Petkakis, and G. Makransky. A study of how immersion and interactivity drive vr learning. *Computers & Education*, 179:104429, 2022.

[243] G. Petri, C. G. von Wangenheim, and A. F. Borgatto. Meega+: an evolution of a model for the evaluation of educational games. *INCoD/GQS*, 3, 2016.

[244] A. Philpot, M. Glancy, P. J. Passmore, A. Wood, and B. Fields. User experience of panoramic video in cave-like and head mounted display viewing conditions. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, pages 65–75, 2017.

[245] P. R. Pintrich. A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of educational Psychology*, 95(4):667, 2003.

[246] J. Pirker, I. Lesjak, M. Parger, and C. Gütl. An educational physics laboratory in mobile versus room scale virtual reality-a comparative study.

In *Online engineering & internet of things*, pages 1029–1043. Springer, 2018.

[247] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 36–39. IEEE, 2017.

[248] W. T. Pouw, T. Van Gog, and F. Paas. An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, 26(1):51–72, 2014.

[249] J. Preece. Citizen science: New research challenges for human–computer interaction. *International Journal of Human-Computer Interaction*, 32(8):585–612, 2016.

[250] F. Pulvermüller. Words in the brain's language. *Behavioral and brain sciences*, 22(2):253–279, 1999.

[251] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, 2020.

[252] H. M. Rao, R. Khanna, D. J. Zielinski, Y. Lu, J. M. Clements, N. D. Potter, M. A. Sommer, R. Kopper, and L. G. Appelbaum. Sensorimotor learning during a marksmanship task in immersive virtual reality. *Frontiers in psychology*, 9:58, 2018.

[253] J. Ratcliffe, N. Ballou, and L. Tokarchuk. Actions, not gestures: contextualising embodied controller interactions in immersive virtual reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2021.

[254] J. Ratcliffe, F. Soave, N. Bryan-Kinns, L. Tokarchuk, and I. Farkhatdinov. Extended reality (xr) remote research: a survey of drawbacks and opportunities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.

[255] J. Ratcliffe and L. Tokarchuk. Evidence for embodied cognition in immersive virtual environments using a second language learning environment. 2020.

[256] J. Ratcliffe and L. Tokarchuk. Evidence for embodied cognition in immersive virtual environments using a second language learning environment. In *2020 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2020.

[257] J. Ratcliffe and L. Tokarchuk. Presence, embodied interaction and motivation: Distinct learning phenomena in an immersive virtual environment. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3661–3668, 2020.

[258] J. Ratcliffe and L. Tokarchuk. Presence, embodied interaction and motivation: Distinct learning phenomena in an immersive virtual environment. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1–8, 2020.

[259] J. Ratcliffe and L. Tokarchuk. Sensorimotor learning in immersive virtual reality: a scoping literature review. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 276–286. IEEE, 2021.

[260] P. A. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt. What is xr? towards a framework for augmented and virtual reality. *Computers in Human Behavior*, 133:107289, 2022.

[261] M. Rauterberg. From gesture to action: Natural user interfaces. *Technical University of Eindhoven, Mens-Machine Interactive, Diesrede*, pages 15–25, 1999.

[262] U.-D. Reips. The web experiment method: Advantages, disadvantages, and solutions. In *Psychological experiments on the Internet*, pages 89–117. Elsevier, 2000.

[263] J. C. Reneker, W. C. Pannell, R. M. Babl, Y. Zhang, S. T. Lirette, F. Adah, and M. R. Reneker. Virtual immersive sensorimotor training (vist) in collegiate soccer athletes: A quasi-experimental study. *Heliyon*, 6(7):e04527, 2020.

[264] C. Repetto, E. Pedroli, and M. Macedonia. Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in psychology*, 8:2136, 2017.

[265] J. C. Richards and T. S. Rodgers. *Approaches and methods in language teaching*. Cambridge university press, 2014.

[266] Y. Rogers and P. Marshall. Research in the wild. *Synthesis Lectures on Human-Centered Informatics*, 10(3):i–97, 2017.

[267] S. Rott. The effect of exposure frequency on intermediate language learners'incidental vocabulary acquisition and retention through reading. *Studies in second language acquisition*, 21(4):589–619, 1999.

[268] R. M. Ryan. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, 43(3):450, 1982.

[269] R. S. Ryan, M. Wilde, and S. Crist. Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. *Computers in Human Behavior*, 29(4):1295 – 1301, 2013.

[270] P. Saalfeld, A. Albrecht, W. D'Hanis, H.-J. Rothkötter, and B. Preim. Learning hand anatomy with sense of embodiment. In *VCBM*, pages 43–47, 2020.

[271] D. Saffo, C. Yildirim, S. Di Bartolomeo, and C. Dunne. Crowdsourcing virtual reality experiments using vrchat. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[272] K. Sakreida, C. Scorolli, M. M. Menz, S. Heim, A. M. Borghi, and F. Binkofski. Are abstract action words embodied? an fmri investigation at the interface between language and motor cognition. *Frontiers in human neuroscience*, 7:125, 2013.

[273] E. Saltz and S. Donnenwerth-Nolan. Does motoric imagery facilitate memory for sentences? a selective interference test. *Journal of Verbal Learning and Verbal Behavior*, 20(3):322–332, 1981.

[274] M. C. Salzman, C. Dede, R. B. Loftin, and J. Chen. A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators & Virtual Environments*, 8(3):293–316, 1999.

[275] F. A. Sanz, A.-H. Olivier, G. Bruder, J. Pettré, and A. Lécuyer. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *2015 IEEE Virtual Reality (VR)*, pages 75–80. IEEE, 2015.

[276] S. J. Savignon. Communicative competence: An experiment in foreign-language teaching. language and the teacher: A series in applied linguistics, volume 12. 1972.

[277] S. J. Savignon. Communicative competence. *The TESOL encyclopedia of English language teaching*, pages 1–7, 2018.

[278] E. Scalona, J. Taborri, D. R. Hayes, Z. Del Prete, S. Rossi, and E. Palermo. Is the neuromuscular organization of throwing unchanged in virtual reality? implications for upper limb rehabilitation. *Electronics*, 8(12):1495, 2019.

[279] M. Schmidt-Kassow, M. Deusser, C. Thiel, S. Otterbein, C. Montag, M. Reuter, W. Banzer, and J. Kaiser. Physical exercise during encoding improves vocabulary learning in young female adults: a neuroendocrinological study. *PloS one*, 8(5):e64172, 2013.

[280] C. Schrader and T. J. Bastiaens. The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, 28(2):648–658, 2012.

[281] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.

[282] T. W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie*, 15(2):69–71, 2003.

[283] N. S. Schutte and E. J. Stilinović. Facilitating empathy through virtual reality. *Motivation and emotion*, 41(6):708–712, 2017.

[284] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[285] K. Y. Segovia and J. N. Bailenson. Virtually true: Children's acquisition of false memories in virtual reality. *Media Psychology*, 12(4):371–393, 2009.

[286] C. Senior, M. L. Phillips, J. Barnes, and A. David. An investigation into the perception of dominance from schematic faces: A study using the world-wide web. *Behavior Research Methods, Instruments, & Computers*, 31(2):341–346, 1999.

[287] J. H. Seo, B. M. Smith, M. Cook, E. Malone, M. Pine, S. Leal, Z. Bai, and J. Suh. Anatomy builder vr: Applying a constructive learning method in the virtual reality canine skeletal system. In *International Conference on Applied Human Factors and Ergonomics*, pages 245–252. Springer, 2017.

[288] M. Shakroum, K. W. Wong, and C. C. Fung. The influence of gesture-based learning system (gbls) on learning outcomes. *Computers & Education*, 117:75–101, 2018.

[289] L. Shapiro. *Embodied cognition.* Routledge, 2019.

[290] R. Shewaga, A. Uribe-Quevedo, B. Kapralos, and F. Alam. A comparison of seated and room-scale virtual reality in a serious game for epidural preparation. *IEEE transactions on emerging topics in computing*, 2017.

[291] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (gsr) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*, pages 2651–2656, 2007.

[292] D. M. Sidhu and P. M. Pexman. Is moving more memorable than proving? effects of embodiment and imagined enactment on verb memory. *Frontiers in psychology*, 7:1010, 2016.

[293] R. Skarbez. Plausibility illusion in virtual environments. 2016.

[294] M. Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4):484–493, 2004.

[295] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[296] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives. Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience*, 2:6, 2008.

[297] M. Slater and M. V. Sanchez-Vives. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3:74, 2016.

[298] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke. First person experience of body transfer in virtual reality. *PloS one*, 5(5):e10564, 2010.

[299] M. Slater and M. Usoh. Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 2(3):221–233, 1993.

[300] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.

[301] F. Soave, N. Bryan-Kinns, and I. Farkhatdinov. A preliminary study on full-body haptic stimulation on modulating self-motion perception in virtual reality. In L. T. De Paolis and P. Bourdot, editors, *Augmented Reality, Virtual Reality, and Computer Graphics*, pages 461–469, Cham, 2020. Springer International Publishing.

[302] J. Sprouse. A validation of amazon mechanical turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods*, 43(1):155–167, 2011.

[303] K. Stanney, C. Fidopiastis, and L. Foster. Virtual reality is sexist: but it does not have to be. *Frontiers in Robotics and AI*, 7:4, 2020.

[304] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp. An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1406–1414, 2016.

[305] A. Steed, F. R. Ortega, A. S. Williams, E. Kruijff, W. Stuerzlinger, A. U. Batmaz, A. S. Won, E. S. Rosenberg, A. L. Simeone, and A. Hayes. Evaluating immersive experiences during covid-19 and beyond. *interactions*, 27(4):62–67, 2020.

[306] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pages 67–76. IEEE, 2016.

[307] L. A. Stein. Imagination and situated cognition. *Journal of Experimental & Theoretical Artificial Intelligence*, 6(4):393–407, 1994.

[308] S. E. Stern and J. E. Faber. The lost e-mail method: Milgram's lost-letter technique in the age of the internet. *Behavior Research Methods, Instruments, & Computers*, 29(2):260–263, 1997.

[309] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93, 1992.

[310] A. Suh and J. Prophet. The state of immersive technology research: A literature analysis. *Computers in Human Behavior*, 86:77–90, 2018.

[311] H. Sun and Y. Gao. Impact of an active educational video game on children's motivation, science knowledge, and physical activity. *Journal of Sport and Health Science*, 5(2):239–245, 2016.

[312] M. Suznjevic, M. Mandurov, and M. Matijasevic. Performance and qoe assessment of htc vive and oculus rift for pick-and-place tasks in vr. In *2017 Ninth international conference on quality of multimedia experience (QoMEX)*, pages 1–3. IEEE, 2017.

[313] J. Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[314] J. Sweller. Cognitive load theory: Recent theoretical advances. *Cognitive Load Theory: Recent Theoretical Advances*, 1(2):29–47, 2010.

[315] L. Talmy. Toward a cognitive semantics - volume 1: Concept structuring systems, 2000.

[316] K. Tamaddon and D. Stiefs. Embodied experiment of levitation in microgravity in a simulated virtual reality environment for science learning. In *2017 IEEE virtual reality workshop on K-12 embodied learning through virtual & augmented reality (KELVAR)*, pages 1–5. IEEE, 2017.

[317] H. Tankovska. Ar/vr headset shipments worldwide 2020-2025, Aug 2020.

[318] K. Tcha-Tokey, E. Loup-Escande, O. Christmann, and S. Richir. Effects on user experience in an edutainment virtual environment: comparison between cave and hmd. In *Proceedings of the European Conference on Cognitive Ergonomics 2017*, pages 1–8, 2017.

[319] M. Tellier. The effect of gestures on second language memorisation by young children. *Gesture*, 8(2):219–235, 2008.

[320] T. D. Terrell. The role of grammar instruction in a communicative approach. *The modern language journal*, 75(1):52–63, 1991.

[321] R. L. Thompson, D. P. Vinson, B. Woll, and G. Vigliocco. The road to language learning is iconic: Evidence from british sign language. *Psychological science*, 23(12):1443–1448, 2012.

[322] J. Tirp, C. Steingröver, N. Wattie, J. Baker, and J. Schorer. Virtual realities as optimal learning environments in sport-a transfer study of virtual and real dart throwing. *Psychological Test and Assessment Modeling*, 57(1):57, 2015.

[323] M. Tucker and R. Ellis. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, 24(3):830, 1998.

[324] T. Tullis, S. Fleischman, M. McNulty, C. Cianchette, and M. Bergel. An empirical comparison of lab and remote usability testing of web sites. In *Usability Professionals Association Conference*, 2002.

[325] E. Tulving. Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 302(1110):361–371, 1983.

[326] H. Tüzün, M. Yılmaz-Soylu, T. Karakuş, Y. İnal, and G. Kızılkaya. The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education*, 52(1):68–77, 2009.

[327] K. Van den Branden. Task-based language teaching. *The Routledge handbook of English language teaching*, pages 238–251, 2016.

[328] S. Van Der Land, A. P. Schouten, F. Feldberg, B. Van Den Hooff, and M. Huysman. Lost in space? cognitive fit and cognitive load in 3d virtual environments. *Computers in Human Behavior*, 29(3):1054–1064, 2013.

[329] F. J. Varela, E. Thompson, and E. Rosch. *The embodied mind: Cognitive science and human experience*. MIT press, 2016.

[330] J. Vasconcelos-Raposo, M. Bessa, M. Melo, L. Barbosa, R. Rodrigues, C. M. Teixeira, L. Cabral, and A. A. Sousa. Adaptation and validation of the igroup presence questionnaire (ipq) in a portuguese sample. *Presence: Teleoperators and virtual environments*, 25(3):191–203, 2016.

[331] M. Vasser and J. Aru. Guidelines for immersive virtual reality in psychological research. *Current Opinion in Psychology*, 2020.

[332] C. Vázquez, L. Xia, T. Aikawa, and P. Maes. Words in motion: Kinesthetic language learning in virtual reality. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pages 272–276. IEEE, 2018.

[333] E. M. Wakefield, C. Hall, K. H. James, and S. Goldin-Meadow. Representational gesture as a tool for promoting verb learning in young children. Boston University Conference on Language Development, Boston, MA, 2017.

[334] E. M. Wakefield, C. Hall, K. H. James, and S. Goldin-Meadow. Gesture for generalization: gesture facilitates flexible learning of words for actions on objects. *Developmental science*, 21(5):e12656, 2018.

[335] D. Waller, E. Bachmann, E. Hodgson, and A. C. Beall. The hive: A huge immersive virtual environment for research in spatial cognition. *Behavior Research Methods*, 39(4):835–843, 2007.

[336] K. R. Walsh and S. D. Pawlowski. Virtual reality: A technology in need of is research. *Communications of the Association for Information Systems*, 8(1):20, 2002.

[337] V. R. Warriar, J. R. Woodward, and L. Tokarchuk. Modelling player preferences in ar mobile games. In *2019 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2019.

[338] J. A. Waterworth, E. L. Waterworth, G. Riva, and F. Mantovani. Presence: Form, content and consciousness. In *Immersed in Media*, pages 35–58. Springer, 2015.

[339] R. P. Weibel, J. Grübel, H. Zhao, T. Thrash, D. Meloni, C. Hölscher, and V. R. Schinazi. Virtual reality experiments with physiological measures. *JoVE (Journal of Visualized Experiments)*, (138):e58318, 2018.

[340] A. D. Wilson and S. Golonka. Embodied cognition is not what you think it is. *Frontiers in psychology*, 4:58, 2013.

[341] C. J. Wilson and A. Soranzo. The use of virtual reality in psychology: a case study in visual perception. *Computational and mathematical methods in medicine*, 2015, 2015.

[342] M. Wilson. The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, 8(1):44–57, 2001.

[343] M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.

[344] R. A. Wilson and L. Foglia. Embodied cognition. 2011.

[345] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[346] A. S. Won, J. Bailenson, J. Lee, and J. Lanier. Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20(3):241–259, 2015.

[347] X. Xiao, S. Zhao, Y. Meng, L. Soghier, X. Zhang, and J. Hahn. A physics-based virtual reality simulation framework for neonatal endotracheal intubation. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 557–565. IEEE, 2020.

[348] Y. Xiao and M. Watson. Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39(1):93–112, 2019.

[349] X. Xu and F. Ke. Embodied interaction: Learning chinese characters through body movements. 2020.

[350] D.-F. Yap, W.-C. So, J.-M. Melvin Yap, Y.-Q. Tan, and R.-L. S. Teoh. Iconic gestures prime words. *Cognitive science*, 35(1):171–183, 2011.

[351] Y. Yuan and A. Steed. Is the rubber hand illusion induced by immersive virtual reality? In *2010 IEEE Virtual Reality Conference (VR)*, pages 95–102. IEEE, 2010.

[352] P. Zahorik and R. L. Jenison. Presence as being-in-the-world. *Presence*, 7(1):78–89, 1998.

[353] L. Zhang and P. G. Healey. Human, chameleon or nodding dog? In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 428–436, 2018.

[354] T. Zindulka, M. Bachynskyi, and J. Müller. Performance and experience of throwing in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.

[355] B. Zinober and M. Martlew. Developmental changes in four types of gesture in relation to acts and vocalizations from 10 to 21 months. *British Journal of Developmental Psychology*, 3(3):293–306, 1985.