ADAPTIVE GUIDANCE IN EXTENDED REALITY LEARNING ENVIRONMENTS

Maheshya Weerasinghe

A thesis submitted for the degree of PhD
at the
University of Primorska
&
University of St Andrews

2023

Please use this identifier to cite or link to this item:
DOI: https://doi.org/10.17630/sta/662

UNIVERSITY OF PRIMORSKA
FACULTY OF MATHEMATICS, NATURAL SCIENCES
AND INFORMATION TECHNOLOGIES
AND
UNIVERSITY OF ST ANDREWS

DOCTORAL DISSERTATION

(DOKTORSKA DISERTACIJA)

ADAPTIVE GUIDANCE IN EXTENDED REALITY LEARNING ENVIRONMENTS

(PRILAGODLJIVO VODENJE V UČNIH OKOLJIH Z RAZŠIRJENO RESNIČNOSTJO)

MAHESHYA WEERASINGHE

KOPER, 2023

Abstract

Learning depends on the dynamics of one's personal circumstances and immediate environment that provides hands- experience. As a result, educators are constantly striving to create personalised learning experiences for learners. The increasing use of technology in education has led to the development of various e-learning systems. However, these systems are limited by their inability to create immersive and interactive learning environments that cater to each learner's individual needs and preferences. Extended Reality (XR) technologies such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) offer a new way of delivering Experiential Learning (ExL) that can meet these challenges. However, existing XR-based learning systems lack the ability to adapt to learners' individual needs and preferences, which may reduce their learning performance. Nevertheless, there is a lack of research and guidance on effectively incorporating XR technologies to design adaptive experiential learning systems. Thus, this thesis aims to contribute new knowledge on how XR technologies can be used to design and develop interactive, adaptive ExL systems that can be integrated into future learning environments. This is accomplished by (i) presenting a comprehensive design space grounded in XR technology and the theoretical underpinnings of learning and instructional guidance, and by (ii) conducting three different user studies, each focusing on an interactive experiential learning system developed based on a particular configuration of the presented design space.

In the first study, the focus is placed on how different representation methods of the future building (paper, desktop and VR HMD) would affect the user experience, dimensions of user engagement, the understanding of the space with minimum guidance, and support users to project themselves into the future office space. The second study explores how different factors of instructional guidance – i.e., the amount of guidance (fixed vs. adaptive-amount) and the type of guidance (fixed vs. adaptive-associations) – would affect the user experience, engagement and the learning outcomes of a language learning scenario. The final study further looks into detail at how different interfaces (AR vs. non-AR) and types of guidance (keyword only vs. keyword + visualisation) would affect user experience, engagement and consequently the learning performances in vocabulary learning.

The results of this research will provide insights into the design and development of interactive XR based experiential learning systems that can meet the diverse learning needs and preferences of individual learners, leading to improved learning outcomes. This work will be useful and of interest to researchers and practitioners who conduct research within the fields of Human-Computer Interaction (HCI), instructional design or education.

Acknowledgements

This thesis represents the journey of the last couple of years of my life, and it would not exist were it not for the guidance, help and support of many people.

First and foremost I would like to extend my sincerest thanks to all my supervisors, Assoc. Prof. Matjaž Kljun, Assoc. Prof. Klen Čopič Pucihar, Prof. Aaron Quigley, Assoc. Prof. Alice Toniolo and Assoc. Prof. Angela Miguel; for their compassion and understanding through all my ups and downs. For all these years they always guided my path patiently, pushing me to be the best version of my research self. Without them all, some of my biggest achievements would never have happened. Thank you.

Special thanks are given to Innorenew CoE (Slovenia), Prof. Jens Grubert and Verena Biener from Mixed Reality Lab (Germany) for their great collaborations. I would also like to acknowledge all of the administrative support, financial and other, of the Faculty of Mathematics, Natural Sciences and Information Technologies (FAMNIT) of the University of Primorska, and Computer Science Department of the University of St.Andrews. Without this support, it would not have been possible to conduct, publish or present this work.

Big thanks to everyone from HICUP Lab for being a part of my life and making me feel at home. Cuauhtli and Julie for helping me on "*Interacting*" with "*Humans*", taking me away from "*Computers*". Nuwan and his wife Sureni for making sure that I stay healthy and don't "*Stay Hungry*", even when I insist to "*Stay Foolish*". Jordan our social butterfly, for finding me users for my user studies and making me laugh always. Elham and Behruz for always being there for me and encouraging me. Each member of HICUP family has played a unique and important role within this chapter of my life. Also, I would like to thank my beloved mathematicians Andrés, René, Blas, and Greg, especially Nina for being such a good listener and a kind friend. In years to come, when I think back on this time, I will remember the memories I made, thanks to all of you. Very special thank to my landlord, Divjak Boštjan for being an amazing human being and providing me with a roof throughout all these years.

I also wish to thank my best friend Kalpi and her family. Thank you for always being there; for never giving up on me; for picking me up at my lowest points; for doing everything you can to help; for always trying to understand, even when I can't understand myself. Thank you for being the best sister and friend I could ever ask for.

Last, and above all, thank you mum for your endless love and for the countless sacrifices you made. Even though now you are not here with me anymore, I know that you will be watching over me forever, from a far away place. Thank you for everything.

Funding
This research was supported by European Commission through the InnoRenew CoE project [Grant Agreement 739574] under the Horizon2020 Widespread-Teaming program and; the University of St Andrews PhD school Scholarship (School of Computer Science). I also acknowledge support from the Slovenian research agency ARRS [program no. BI-DE/20-21-002, P1-0383, J1-9186, J1-1715, J5-1796, and J1-1692].

Candidate's declaration

I, Anuradhi Maheshya Weerasinghe, do hereby certify that this thesis, submitted for the degree of PhD which is approximately 22,100 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews and the University of Primorska in September 2018 as part of a joint programme.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Signature of candidate Date 2023.11.05

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature of supervisor Date

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Anuradhi Maheshya Weerasinghe, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Signature of candidate Date 2023.11.05

Signature of supervisor Date

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Anuradhi Maheshya Weerasinghe, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Signature of candidate Date 2023.11.05

Contents

List of Figures

List of Tables

To my deeply missed mum,
my best friend Kalpi,
HICUP Lab
and “Confucius”....

Chapter 1

Introduction

Learning is instinctive. It is a conscious activity in formal education and training, but it also lives in one's subconscious, thinking, reflection and even imagination. Learning depends on the dynamics of personal circumstances and the immediate environment. Students can learn best when they are immersed in learning experiences that are active, stimulating, meaningful and interactive. These can be small team projects, study trips, internship programmes or any other experiential learning task. This means learning can occur not only in classrooms but anywhere at any time in our daily lives. The advancement of platforms such as *Metaverse* [1] and *NVIDIA Omnivers* [2] open up a world of possibilities for creating engaging and interactive Extended Reality (XR) – mixing real world with digital content or replacing it altogether – learning environments, as they can better support different forms of experiential learning and group collaboration not always possible to be achieved in traditional classroom or with existing learning tools. Research in Human-Computer Interaction (HCI) aims to investigate how computer powered technologies can be utilised to support humans as they interact with them. This work, takes an HCI perspective in order to better understand how novel technologies, such as XR, can be used to design and develop interactive Experiential Learning (ExL) systems that can be integrated into future classrooms as well as our day-to-day lives.

1.1 Research Context

Experiential Learning (ExL) is a well-known learning approach used in education, training, facilitation, coaching and organisational development [81, 71, 44]. ExL refers to the process of learning through experience or "learning by doing" and can be described by one of the most influential models in ExL [81] – Kolb's learning model [69]. This model defines ExL as "the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience" [69]. The model is represented by four cyclical learning stages visible also in Figure 2.1:

1. Concrete experience (feeling): doing or having a novel experience.

[1] https://about.meta.com/metaverse/
[2] https://www.nvidia.com/omniverse/

2. Reflective observation (watching): reviewing and reflecting on the novel experience, focusing on whether there were any discrepancies between the experience and understanding.

3. Abstract conceptualisation (thinking): what was learned from the experience, thinking how to improve for the next time.

4. Active experimentation (doing): planning or practically applying what has been learned.

It is possible to enter the learning cycle at any stage, but all stages in the cycle must be addressed for a meaningful learning to occur [69].

The degree to which these stages occur spontaneously is a point of departure for different approaches. Some researchers promote an unguided approach in which the learner freely explores and constructs meaning through an unstructured process [25, 47, 21]. At the core of these approaches is the belief that learners learn by being immersed in an authentic environment. As they make sense of it and build their own cognitive structures, they develop a deep understanding. Some researchers on the other hand promote the creation of guided experiences that intentionally structure the environment and/or sequences of events based on learner's personality, creativity, motivation, attitude, etc., including prior knowledge and previous experiences [67].

In learning, including ExL, guidance is beneficial. The term guidance has a broad meaning, but we adopt the generally understood definition of guidance as "any form of assistance offered to users so they can achieve a learning goal". In more precise terms, the role of guidance is "to simplify, provide a view on, elicit, supplant, or prescribe the scientific reasoning skills" [79] involved in pursuing a goal. It can be explicit, as when a teacher provides instructions on how to solve a mathematical problem with all the in-depth explanations of concepts and skills students need to learn, or provides feedback on the current performance. Or it can be implicit, as it is provided within the environment for users to find, interpret and use in order to progress in the task at hand [10, 26]. In either case, the amount, the type and timing of guidance play an essential role in engagement and consequently, in the learning outcomes [27, 128, 63].

While several educational definitions of guidance exist [32, 79, 3], in the context of this thesis we focus on instructional "guidance", which is defined as "providing users with accurate and complete procedural information (and related declarative knowledge) that they have not yet learned in a demonstration about how to perform the necessary sequence of actions and make the necessary decisions to accomplish a learning task" [32]. Guidance also has different dimensions such as "how much" (amount) instructional support is given [86, 126, 139], "what kind" (type) of support is given [3], and "when" (timing) is the support given [9, 85, 115].

The term Extended Reality (XR) describes interactive environments generated by computer technology combining real and virtual worlds. XR is an encompassing term for Augmented Reality (AR), Mixed Reality (MR) and Virtual Reality (VR). It describes interactive environments generated by computer technology combining real and/or virtual worlds. XR can provide an environment for simulating various experiences by combining the physical environments with adaptive digital elements that can be manipulated by the user. XR can thus replicate the real-world ExL scenarios and provide instructional guidance by showing information in a coherent and meaningful way.

Researchers have previously leveraged XR systems as tools for guiding or teaching different skills that can be learned best through experience [128, 124, 17, 51, 137, 119, 8, 76]. As the review of recent XR studies highlight (Chapter 2 and Chapter 3), the majority of studies have primarily focused only on learning stages of Kolb's ExL cycle [128, 17, 119, 8, 76]. However, according to Kolb, the learner must continue cycling through all the four stages of the ExL cycle, thus creating a "learning spiral of ever-increasing complexity" in order to gain a better learning outcome [70]. Currently, we are unaware of any literature that focuses on providing guidance to support learners move from one stage to another in the ExL cycle and progress on the learning spiral.

Moreover, these approaches have focused only on providing generic guidance where all users receive the same set of general instructional cues, with the same amount, in the same way [128, 17, 60, 119, 76]. However, the theories and other groundwork on learning and instructional guidance claim that tailoring guidance to the different needs of users (changing the type, the amount and timing) plays a critical role in learning as it affects the engagement and consequently the performance of individual learners differently [9, 115, 139]. Currently, besides the limited numbers of studies such as [124], there is a lack of research on providing adaptive guidance by adjusting the type, the amount and timing of the instructions for different user needs.

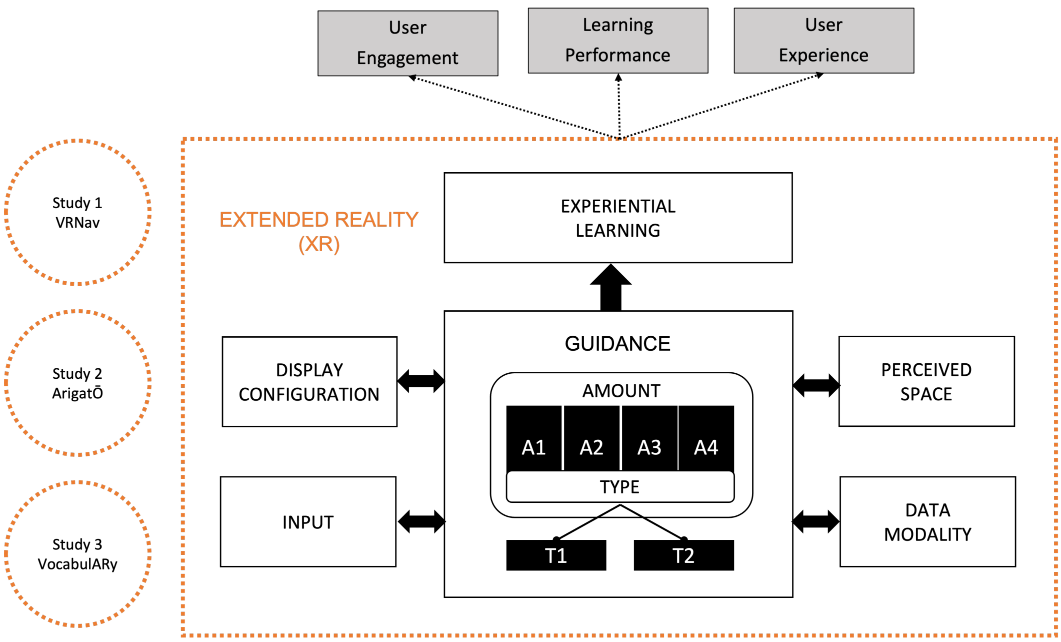


Figure 1.1: Design space overview and research method illustrating the thesis overview. The thesis constitutes of 3 user studies (i.e. VRNav, Arigatō and VocabulARy) exploring the design space of XR experiential learning systems with a focus on guidance amount (A1, A2, A3, A4) and type of guidance (T1, T2). Throughout the thesis, dependant variables cover User Engagement, Learning Performance and User Experience performance metrics. When designing XR learning systems a particular attention was given to Display configuration (e.g. HMD, Desktop), Input system (e.g. controller based, mid-air gesture based), Data presentation modality(e.g. 2D, 3D, text) and Perceived space (e.g. local, remote).

The use of XR technology in the design of interactive ExL environments has the potential to transform the way students learn and interact with educational systems.

However, existing XR-based learning systems are often not fully tailored to the diverse needs of learners, leading to lack of engagement and poor learning performance.

Thus, this thesis aims to explore the *potential of XR technologies in designing and implementing experiential learning environments that cater to the unique needs of individual learners. The goal is to enhance understanding and increase engagement by providing interactive and stimulating environments supported by effective guidance.*

To achieve this, the thesis presents a comprehensive design space that incorporates the characteristics of XR environments and the underlying theories of learning and guidance. The design space considers various dimensions such as display configuration, data modality, input, perceived space, experiential learning, and guidance, which are discussed in Chapter 3.The research employs the design space to (i) systematically explore the existing literature and identify the gaps and (ii) develop interactive experiential learning systems tailored to different learning scenarios –e.g., Study 1:*VRNav* focuses on spatial learning and the effect immersion has on relocating to the new office space while Study 2:*Arigatō* and Study 3:*vocabulARy* focus on second language learning. The primary focus of this thesis is to evaluate the effectiveness of different configurations within the design space for creating these learning experiences in XR environments. The evaluation will encompass various aspects, including user experience, user engagement (covering affective, behavioural, and cognitive factors), and ultimately, learning performance. To better present thesis structure Figure 1.1 shows the overview of the design space and the research methods we follow.

The research questions that will be explored in this thesis are presented in Section 1.2

1.2 Research Questions and Hypotheses

This section lists a set of research questions and hypotheses that we investigate in this thesis.

> **RQ1:** How can XR technology be used to create immersive and interactive experiential learning environments by exploring the importance and effectiveness of different configurations of XR, ExL and instructional guidance?

H1-1: *Constructive experiential learning experiences can be designed by investigating different configurations of XR, ExL and instructional guidance.*

H1-2: *User engagement and learning performance (e.g. user actions, cognitive load, attention, task completion time, and error rate) vary for different configurations of ExL environments.*

> **RQ2:** What are the XR interventions that can be used to remodel the type and the amount of instructional guidance and nudge learners to move from one stage to another in the ExL cycle?

H2-1: *User engagement and learning performance (e.g. user actions, cognitive load, attention, task completion time, and error rate) can be efficiently measured in XR environments to remodel the instructional guidance.*

H2-2: *XR interventions can be embodied as explicit guidance to support the learner to proceed spontaneously from one stage to another in Kolb's ExL cycle.*

RQ3: How can XR technology be used to create immersive and adaptive experiential learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators (e.g. providing guidance, detecting and deciding on the amount of guidance and selecting the appropriate XR interventions to support guidance [RQ2])?

H3-1: *XR can be used to create adaptive educational systems that would be able to provide immersive and efficient learning environments which outperform standard learning systems.*

RQ4: How do the different configurations of XR (i.e., display configuration, input type, data modality, perceived space) and/or degree of dynamic adaptions of guidance (i.e., amount and type – or what can be jointly called "personalisation" of instructional guidance) affect the user experience, engagement and, consequently, the learning outcomes of immersive ExL environments?

H4-1: *The display configuration, input type, data modality, and perceived space of XR environments would affect the user experience, engagement and learning outcomes of an ExL task.*

H4-2: *The type and the amount of instructional guidance generated by XR environments would affect the user experience, engagement and learning outcomes of an ExL task.*

H4-3: *Adaptive amount of guidance will outperform fixed guidance in subjective and objective performance measures.*

1.3 Methodology and Approach

Due to the interdisciplinary and challenging nature of the research, we conducted three different studies to investigate the above mentioned research questions. The first study (*VRNav*: Chapter 4) utilises an immersive VR environment to investigate, measure and compare how different representation methods – i.e., VR head-mounted display (HMD), desktop VR and paper-based 2D floor plans – of the office building would, (i) affect the user's engagement and the understanding of the space and, (ii) support the users projecting themselves into the building. More specifically, *VRNav* is centred around the first research question (RQ1) and last research question (RQ4), and their corresponding hypotheses presented in the previous section (Section 1.2).

The second study (*Arigatō*: Chapter 5) includes an adaptive guidance AR system for language learning that explores RQ2, RQ3, and RQ4, presented in Section 1.2. In particular, in this study, we aim to investigate how the different factors of instructional guidance, – i.e., the amount (fixed vs. adaptive-amount) and the type (fixed vs. adaptive-associations) of guidance – would affect user engagement and, consequently, the learning outcomes of a language learning scenario.

The final study (*vocabulARy*: Chapter 6) includes an AR system for vocabulary learning that investigates RQ2 and RQ4, presented in Section 1.2. In this study, we aim to explore how the interface (AR vs. non-AR) and different types of guidance (keyword only vs. keyword + visualisation) would affect user engagement and, consequently, the learning performances in vocabulary learning. This was achieved by designing a HMD based AR vocabulary system that visually annotates objects in the user's surroundings and comparing it to a non-AR systems on a tablet.

In order to maximise the effectiveness of each study, a mixed methods approach was adopted, and details decided on a study-by-study basis (see Chapter 4, Chapter 5 and Chapter 6 for more details). Factors considered when making these decisions included: study design, number of participants, duration of the study, data collection and research aim of the study. As a result, the research described in this work was conducted using different, but interrelated methodological approaches. Thus,in all studies, the data was collected using both qualitative and quantitative techniques. In addition, the combination of descriptive, observational and experimental approaches allowed for a broad perspective on how participants are influenced by technology as a tool for experiential learning settings. At the same time, this allowed for detailed insights into specific student-centred technology interactions within experiential learning scenarios based on real-word phenomena and applications. Furthermore, each of the studies can be considered individually to create a loose picture of the whole - a useful facet for researchers in HCI and education.

1.3.1 Ethics Approval

All of the studies conducted have been subject to full ethical approval from both the University of Primorska and the University of St.Andrews. In addition, all participants participated on a voluntary basis and consented to their participation in the work. All documents pertaining to ethical considerations of this work are included in Appendix A, Appendix B, and Appendix C.

1.4 Contributions

While HCI is a subsection of Computer Science, it is a interdisciplinary field, which draws on research from many areas, including Psychology, Social Science, Engineering and Design. The research scope of this thesis touches upon, and draws from learning theories and instructional design theories. The findings presented here will primarily be of interest to researchers within the field of HCI and instructional design. However, it may also have be of interest to those within education, and educational technology fields. Through the interdisciplinary work described in this thesis, our work contributes to research in the areas of HCI, instructional design and education. In this research we propose a novel design approach to develop adaptive and interactive XR environments and show how these can be used to improve experiential learning and promote user engagement. In particular, our contributions are outlined more specifically in the following:

- A series of user studies demonstrating the usability of XR for designing interactive, ExL environments (Chapter 4, Chapter 5 and Chapter 6) that maximise user engagement and, consequently, learning performance.

- Successful examples for XR interventions that are effective in designing guidance (Chapter 4, Chapter 5 and Chapter 6) and supporting the learners move from one stage to another in the ExL cycle (Chapter 5).
- Successful examples for developing interactive ExL systems in XR with minimum guidance to maximise user engagement (Chapter 4).
- Design considerations for adapting the type of guidance in immersive ExL environments to maximise user engagement and learning performance (Chapter 6).
- Design considerations for adapting the amount of guidance in immersive ExL environments to maximise user engagement and learning performance (Chapter 5).
- A comprehensive analysis of variables such as user engagement (i.e., affective, behavioural, cognitive), learning outcomes and user experience, to understand the usability of XR for designing interactive, adaptive ExL environments (Chapter 7).
- Design considerations for the development of future interactive, experiential learning systems, consolidated from the results of analytical and empirical research (Chapter 7).

1.5 Thesis Outline

This thesis is organised into seven chapters, a high level overview can be seen in Figure 1.2. The outline threads together the publications discuss in this thesis and combines them into a single unified body of work.

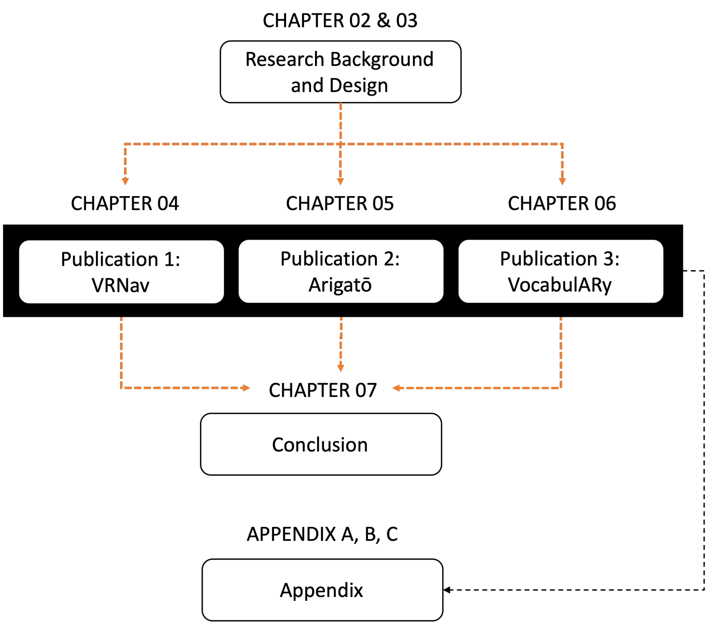


Figure 1.2: A high level overview of the work within and structure of this thesis.

Chapter 2 introduces the background and related research relevant to this work. It provides a sound knowledge of theoretical underpinnings of learning and instructional design for developing learning systems, and XR technology as part of HCI. Additionally, this chapter looks at how AR and VR technologies have been researched before, identifying a research gap regarding the use of technologies as tools for learning in educational environments.

Chapter 3 introduces a broader design space by exploring the domain, and highlighting the importance and effectiveness of different configurations of XR-based experiential learning systems as described in Chapter 2. It also provides a comprehensive understanding of the dimensions of the design space and shows (i) how it was used to systematically explore the existing body of literature and identify knowledge gaps, and (ii) how it was utilised to design and develop various learning scenarios.

Chapter 4 includes the journal article published based on the Study 1: *VRNav*, that was conducted to investigate how different representation methods (VR, paper and desktop) would affect the dimensions of user engagement and the understanding of the space with minimum guidance. Chapter 5 presents the article published based on the Study 2: *Arigatō*, that explores the effects of adaptive guidance on engagement and performance in AR learning environments. Chapter 6 includes the journal article published based on the Study 3: *vocabulARy*, that investigates the effects of different types of guidance on engagement and performance of vocabulary learning in AR.

Chapter 7 concludes the work described in this thesis, identifying the contributions were made. This also establishes the limitations of this work and proposes directions that future research could take.

Appendix A, contains the ethical and study documentation relevant for Study 1:VR-Nav. Some study findings, including statistical analyses, are reported in full here also. Documents regarding Study 2: Arigatō, including ethics and study resources, are included in Appendix B. Appendix C, provides the ethical and study documents required during Study 3: VocabulARy.

Chapter 2

Research Background

"Learning without reflection is a waste. Reflection without learning is dangerous."

Confucius

The scope of the research, which discusses the role of technology in designing adaptive learning systems, draws upon the theoretical underpinnings of learning and instructional guidance and places them within the field of HCI by evaluating technologies used to develop immersive learning experiences. The following sections provide the reader with relevant background information on the educational context of this research, approaches to instructional guidance design, and existing novel technologies for developing immersive learning systems that can be integrated into future learning environments. In this context, Section 2.1 examines how psychologists and educationalists have explained learning from different perspectives, with a particular focus on experiential learning. Section 2.2 introduces the concept of instructional guidance and its broader position in designing adaptive instruction for experiential learning systems. This includes an insight into the debate about the amount, type and timing of instructional guidance that is most effective and efficient for learning and comprehension. Section 2.3 provides a brief description of engagement and its dimensions (i.e., affective, cognitive, behavioural and agentic) that are useful for designing effective learning environments. Before concluding this chapter, Section 2.4 has outlined the novel technologies for developing immersive learning systems and the research that has been conducted to better understand its use in context was explained.

2.1 Learning

Learning is of interest to several research disciplines. Numerous psychologists and educators have explained the concept of learning from different points of view. Some define learning as a process, others as a change in performance, while some define it as the acquisition and retention of knowledge [13, 92, 142, 6]. These are parts of different learning paradigms and theories described in the following sections.

2.1.1 Learning Paradigms and Theories

Learning theories are conceptual frameworks that describe how people acquire, process, and retain knowledge during the learning process. They fall into one of several learning paradigms, including behaviourism, cognitivism, constructivism, and humanism [13].

Different theories are appropriate for different situations and learning outcomes. There is no single accepted definition of learning, since it depends on one's point of view or a learning paradigm. Most commonly accepted learning paradigms suggest that learning is:

- a visible change in one's behaviour, which can be measured [13, 92] (i.e., providing feedback in a game to learners and providing reinforcement to positively impact performance)
- the active process of acquisition (including insight, information processing, memory, perception) of new knowledge and developing adequate mental constructions [13, 142] (i.e., stimulating various regions of the brain and increase the number of consolidation processes through repetition and improve reflexes, promote critical thinking, and help people learn)
- an active, socially enhanced process of knowledge construction based on one's own subjective interpretation of the objective reality [13, 16] (i.e., collaboratively and cooperatively engaging in a task in order to achieve a goal in a game).
- a natural desire of human beings, a mean of self-actualisation and developing personal potentials [13, 6] (i.e., learning in a game through a cycle of concrete experiences, reflective observation, abstract conceptualisation and active experimentation).
- the process of connecting to information sources containing actionable knowledge and maintaining those connections [13, 72] (i.e leveraging game skills that are transferable across media, platforms and tools to expand students' learning networks).

Theories within the same paradigm share the same basic point of view [13]. It has to be stressed that each of the paradigms has attracted both supporters and critics. Presenting all possible views is beyond the scope of this chapter and what follows is a brief overview of the paradigms mentioned above.

Behaviourism states that all behaviours are learned through the interactions with the external environment. Behaviourists do not attempt to analyse its inner processes of mind such as thoughts, feelings, or motivation. From a behaviourist perspective, a learner starts off as a clear state and simply responds to environmental stimuli. These responses can be shaped through positive and negative reinforcement (a reward for desired or a punishment for undesired behaviour), increasing or decreasing the probability of repeating the same behaviour [13, 92]. Sign learning model presents learning as the acquisition of knowledge through meaningful behaviours (Tolman 1922).

Cognitivism is a learning paradigm focused on the inner mental processes of humans: how the human brain perceives things, how it makes memories and create new knowledge [13, 142]. The cognitive approach to learning sees the learner as an active participant in the learning process, acquiring new knowledge and constructing mental

Time Period	1900s	1950s	1960s	1970s	1980s	2000-
Learning Theories	Behaviorism: Based on observable changes in behaviour. Behaviorism focuses on a new behavioural pattern being repeated until it becomes automatic. (***Sign Learning:*** *presents learning as the acquisition of knowledge through meaningful behaviours* [127])					
		Cognitivism: Based on the thought process behind the behaviour. Changes in behaviour are observed, and used as indicators to what is happening inside the learner's mind. (***Brain-Based Learning:*** *presents learning as a cognitive development process which emphasises how people learn differently as they grow, mature socially and emotionally, and cognitively* [61])				
			Humanism: Based on the natural human desires & development of personal potentials. (***Experiential Learning:*** *defines the process of learning as "learning through reflection on doing". According to the Experiential learning theory, knowledge results from the combination of grasping and transforming experience* [81, 71])			
				Constructivism: Based on the premise that we all construct our perspective of the world through, individual experiences and also social interactions. (***Problem-Based Learning:*** *suggests that learning is more effective when learners are faced with a real-life practical problem and empowers learners to conduct research and apply knowledge and skills to develop a viable solution* [114] ***Inquiry-Based Learning:*** *encourages learners to use world connections through exploration and high-level questioning while learning* [78])		
						Connectivism: Based on the process of connection forming in the context of technological development. (***Simulation-Based Learning:*** *provides learners with an experience of working on a simplified simulated world or system* [77])

Table 2.1: A brief overview of learning paradigms including the major theories of learning and their period of development.

constructions based on prior knowledge and experience. Unlike behaviourism, it tries to understand the complex cognitive processes by searching for associations between learning and information processing, perceptions and memory.

Constructivism is a learning paradigm claiming that learners construct their own knowledge of the world through experiencing things and reflecting on those experiences [13, 16]. Constructivism's approach to learning differs from behaviourism and cognitivism in that it perceives learning as an active, socially supported process of knowledge construction. As such, learner constructs their own subjective interpretation and meaning of what is being learnt of objective reality.

Humanism defines learning as a natural human desire, based on self-actualisation and development of personal potentials [13, 6]. It emphasises the importance of every individual in that they are striving towards happiness through self-achievement while being responsible for their own actions. Individuals should also have control over the learning process, which should be based on observing and exploring. The learning process is considered more important than the learning outcomes. Since the control is in the learner's hands the role of the teacher is to encourage, motivate and provide reasons for embarking on the learning journey.

Connectivism claims that learning occurs not only in individual but also within and across networks. As such, learning resides also outside an individual such as within an organisation or web. The connections and the network of an individual are thus more important than their current state of knowledge. Connectivism is proposed as a learning paradigm for the digital age, which attempts to approach learning and knowledge in the context of technological development [13, 72]. Connectivist learners share and communicate dynamic knowledge creation through networked interaction with machines and other people.

Learning theories fall into one or more learning paradigms. Experiential learning theory, for example, is based on the humanistic and constructivist perspectives, which emphasise that learning occurs naturally, and that experience is crucial in knowledge acquisition. Table 2.1 provides an overview of learning paradigms and learning theories that fall predominantly into one paradigm. As our work focuses on designing immersive experiential learning environments that can be integrated into future educational settings, the following section delves into more details on ExL and some of the popular ExL models.

2.1.2 Experiential Learning

Experiential Learning (ExL) theory is based on the idea that learning is a process of acquiring knowledge through experience or "learning through reflection on doing" [69, 81, 65]. It is one of the major theories of education grounded in constructivism and humanism learning paradigms [39, 113]. As Keeton and Tate express it [65], ExL "involves direct encounter with the phenomenon being studied rather than merely thinking about the encounter or only considering the possibility of doing something with it". These thoughts about experiential learning continue to evolve overtime and, several models as "theories-in-use" for ExL have been proposed. To shed further light on our understanding, we can zoom in to examine some of the more recent and popular models related to ExL.

Kolb's Model of Experiential Learning

Kolb's model is one of the most popular models of ExL [81]. As shown in Figure 2.1, Kolb describes ExL as a four-part process, where the learner is asked to engage in a new experience, actively reflects on that experience, conceptualises that experience and integrates it with prior experiences and further, makes decisions based on the created concepts. According to Kolb, the learner must continue cycling through the four-parts, thus creating a "learning spiral of ever-increasing complexity" [81]. A learner might begin anywhere in the cycle at any level of knowledge concerning the subject matter. The facilitator's job is to guide them through each stage in an ever increasing level, expanding their learning of a topic. Finally, through this model Kolb stresses that learning is a process where "ideas are not fixed and immutable elements of thought but are formed and re-formed through experience" [69].

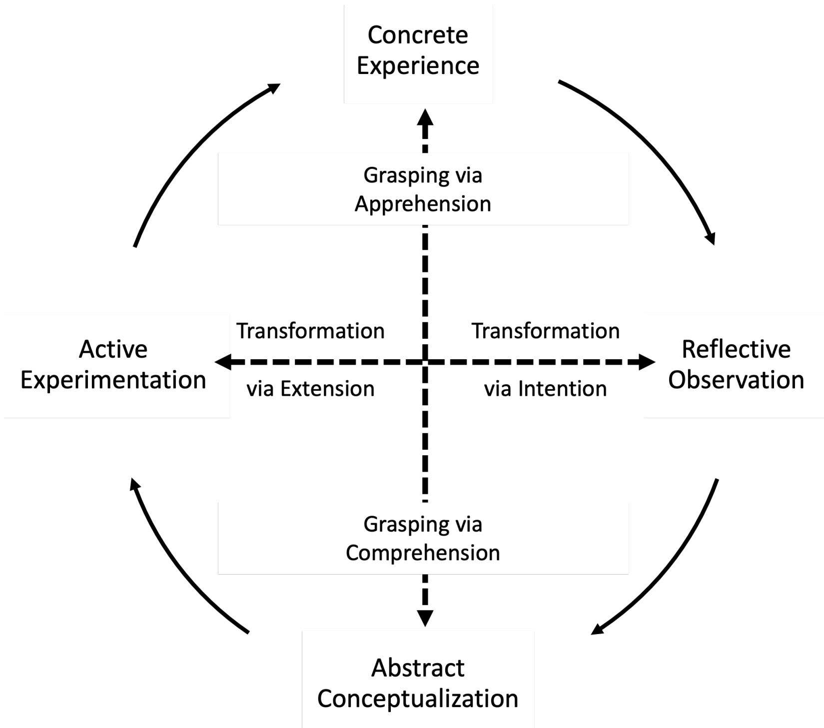


Figure 2.1: Experiential learning cycle by Kolb [69].

Boud and Walker's Stages in Experiential Learning

Boud and Walker discuss ExL as a series of stages where there is some kind of preparation done before a learning event, the actual experience itself, and then reflection to "debrief" the learner on what took place [20]. This idea incorporates two significant aspects of Kolb's model: experience and reflection. It also adds a third aspect: preparation for the event, that the authors think is important in learning. When considering preparation for a learning event, the facilitator needs to focus on what experiences the learners bring and what they want to learn. Overall, through their model Boud and Walker stress that "learners bring with them 'intent', which may or may not be able to be articulated, and which influences their approach to the event" and thus "greater use can be made of learning events if the learners prepare beforehand" [20].

Dean's Process Model of Experiential Learning

Dean's Process Model of ExL is based on stages adapted from Pfeiffer and Jone's work [98]. This model is presented as a series of stages in the process of developing and implementing an ExL activity [38]:

1. Planning - Getting Ready to Start
2. Involvement - Getting Started
3. Internalisation - Learning by Doing
4. Reflection - Making Meaning
5. Generalisation - Making Connections
6. Application - Transfer of Learning
7. Follow-up - Assessment & Planning

As with Boud and Walker [20], Dean also sees ExL as a process that the facilitator goes through to develop the learning experience. The central concept of Dean's model relates to the other models of ExL [69, 20], in that there needs to be some form of *experience* (involvement and internalisation) and a *reflection* on that experience for meaningful learning to occur. Overall, Dean's process model of ExL places an important emphasis on the role of the facilitator. The facilitator will have to assume the leadership role in helping the group of learners to get involved in the learning activity, to process the learning, and to apply the learning to activities on their experiences.

Joplin's Five Stage Model

Joplin's Five Stage Model of ExL also follows the "action-reflection" process [62]. However, Joplin adds three other stages that are similar to Boud and Walker's, and Dean's models [20, 38]. The first stage of Joplin's model is *focus*, which defines the task to be completed and looks at the learners' attention on that task. Second is *action*, where learners must become involved with the subject matter in a physical, mental, or emotional manner. Third and fourth stages are *support* and *feedback*. These are present throughout the learning experience and are provided by the facilitator or fellow learners. The fifth and last stage is *debrief*, where the learners and facilitator sort and order the information and reflect on its implications. Joplin stresses that "experience alone is insufficient to be called experiential education, and it is the reflection process, which turns experience into experiential education" [62].

Amongst all the models for ExL, Kolb's model has been identified as one of the most cited and widely used learning model in the field of education [81, 71]. Besides, after examining the other popular models for ExL, it can be seen that most of the models have adopted Kolb's model as a groundwork for designing their conceptual models [20, 38]. As such, in this thesis, we use Kolb's ExL model as the base model for designing our learning system and further exploring the field.

Learning, including experiential learning, benefits from instructional guidance. However, in a learning environment, different learners may process information most effectively through their own preferred modes of instructions. Therefore, when designing instructional support, instructional designers follow various theories and perspectives that address how people learn and how the cognitive processes behind the learning experience work. The following section provides a background on these theoretical perspectives and their broader position in designing adaptive instruction for experiential learning systems.

2.2 Instructional Guidance

The instructional support provides during learning has been categorised and explained using different concepts and terms, such as above mentioned learning paradigms (i.e., behaviorism, cognitivism and constructivism, etc.) [88, 126] and instructional methods (e.g. lectures, case studies, peer feedback, quizzes, etc.) [35, 125]. Such paradigms and methods often emerge from different learning theories and sometimes various belief systems, and philosophies [35, 88, 52, 139, 86]. Therefore, when trying to employ a construct, such as "instructional guidance", we depend on the fact that the use of different conceptual frameworks and theories tend to define and operationalise instructional support in different ways. However, the main purpose of instructional guidance is to reduce the learner's potential cognitive overload (i.e., the amount of information that working memory can hold at one time) by providing appropriate information in the right amount, at the right time and in the suitable format. This section provides an insight into the debate about the amount, type and timing of instructional guidance that is most effective and efficient for learning and comprehension (the transfer of knowledge).

Amount of Guidance

Debates about the impact of instructional guidance and "how much" instructional support needs to be provided in a learning environment have been ongoing for at least the past century [86, 67, 75, 126]. For example, is it better to tell learners what they need to know by presenting them the essential information, or is it better to allow learners to discover or construct essential knowledge for themselves? Koedinger and Aleven called this issue the "assistance dilemma" [68], i.e., deciding whether to provide or withhold assistance. The contrast between the two practices can be better understood as a continuum. On one side of this continuum are those supporting the hypothesis that people learn best in an unguided or minimally guided environment [25, 118]. On the other side are those suggesting that learners should be given a direct instructional guidance on the concepts and procedures required by a particular task [35, 86, 67]. In this case, the "direct instructional guidance" is defined as "providing information that fully explains the concepts and procedures that learners are required to learn as well as learning strategy support that is compatible with human cognitive architecture" [67]. Learning, in turn, is defined as a change in human long-term memory.

The minimally guided approach has been explained using various learning theories including problem-based learning [114], inquiry learning [79] and ExL [70]. There seem to be two main assumptions underlying instructional support approaches using minimal guidance. First, they challenge learners to solve "authentic" problems or acquire advanced knowledge by assuming that letting learners construct their own knowledge leads to the most effective learning experience. Second, they appear to presume that knowledge can best be acquired through experience based on the procedures of the task. In other words, in a minimally guided learning environment, students are seen as active learners and are given opportunities to digest contents for themselves rather than as passive learners who merely follow instructions. These guided approaches have been widely accepted as major teaching methods by teachers and educators with constructivist views of learning [138].

In contrast, there also have been decades of efforts to discourage educators from adopting minimally guided learning approaches. The past half-century of empirical research on the topic has provided evidence that minimal guidance during instruction is significantly less effective and less efficient than guidance specifically designed to support the cognitive processing necessary for learning [67]. The authors also highlight evidence for the superiority of fully guided instruction – i.e., provide a direct instructional guidance on the concepts and procedures required by a particular task– , which is explained in the context of human cognitive architecture, expert-novice differences, and cognitive load.

Furthermore, while all these disputes agree that the amount of guidance, i.e., "all-or-none" is important to think about, Wise and O'Neill, present a review of evidence to support that the optimal amount of guidance often is an intermediate amount and the *granularity* of the advice provided in a task (i.e., the level of details) is equally important [139]. To illustrate this, Wise and O'Neill revisit one of the primary sources of evidence cited by Kirschner et al., "the worked-example effect", for their claim that more guidance is always better [29, 96, 129, 90]. A worked-example is a step-by-step demonstration of how to perform a task or how to solve a problem [33]. Beyond the worked-example literature, Wise and O'Neill also explain other studies which suggest that, aiming for the right level of granularity in guidance is a better guideline than "more is always better" – e.g., Nadolski, Kirschner, and van Merriënboer's work explains that breaking the task of preparing a legal plea into four steps was more effective in supporting their target population of learners than presenting the task as a whole (one step) or in nine steps [94].

Type of Guidance

Research on memory and learning has shown that comprehension and recall depend on different types of instructional methods and techniques that can be used to process and store information [40]. Mnemonic is one type of instructional technique designed for enhancing the memory and recall [141, 101, 84, 99]. This technique connects new learning to prior knowledge through the use of visual and/or acoustic cues. The basic types of mnemonic techniques rely on the use of phonetic systems, key words, rhyming words, or acronyms [101, 84]. Table 2.2 provides short descriptions of some major mnemonic techniques.

Each mnemonic is designed to help remember a specific kind of information. "Acronyms and acrostics", for example, help a user remember word lists, but the words can refer to anything (e.g., the rainbow colours, or the planets) perhaps explaining why learner use first-letter strategies more than other mnemonics [117]. Unfortunately, there is mixed evidence about whether first-letter mnemonics actually facilitate recall. Researchers have argued that first-letter mnemonics are not effective retrieval cues, and thus will likely not aid recall unless learners are

Mnemonic	Description
Link method	Interactive visual imagery connects items in a list, making a chain. Item 1 is joined with item 2; a separate image joins item 2 with item 3 and so on. Thus, retrieving one item in the list cues the next item.
Method of loci	First, a memory palace—a mental map of a building that one know well, such as a house—is memorised. Then, imagery is used to store list items at different locations throughout the memory palace and items are retrieved by "walking" through the palace.
Peg system	A "peg list", or a list of concrete objects in a specific order (e.g., one is a bun, two is a shoe, three is a flea) is learned. Then, visual imagery combines the to be remembered items with the peg items. Items can be retrieved by thinking of a number and the corresponding peg, which cues the target item.
Keyword method	First, a keyword is found that sounds like the unfamiliar word (e.g., "dentist" sounds like "la dent"). Then imagery joins the keyword with the definition of the unfamiliar word (an image of a "dentist" holding a large "tooth"). Seeing "la dent" activates dentist, which in turn should activate tooth.
Phonetic system	Each number corresponds to a consonant sound ($1 = t$, $2 = n$, $3 = m$ etc.). Then numbers can be remembered as words, using vowels as necessary. For example, 321 can be remembered as "manatee". Words can be decoded back into numbers.
Acronyms	The first letters of a list of words are used to create a new word. For example, the colours of the rainbow (red, orange, yellow, green, blue, indigo, violet) can be remembered as **ROYGBIV**. Each letter serves as a retrieval cue for the target items.
Acrostics	The first letters in a list of words serve as the first letters in a new sentence or phrase. For example, the colours of the rainbow can be remembered as **R**ichard **O**f **Y**ork **G**ave **B**attle **I**n **V**ain. The first letter in each word of the acrostic serves as a retrieval cue.

Table 2.2: Descriptions of popular mnemonic techniques and systems [101].

already familiar with the material [28]. Thus, despite being theoretically applicable to a wide range of educational materials and popular among learners, first-letter mnemonics may not be effective memory aids.

In contrast, in the field of language learning, the "keyword method" [14] or creating mental associations to a known language has proven to be effective in the memorisation of vocabulary. In the keyword method, students associate the sound of a word they want to learn to one they already know in either their first language or the target language. They then mentally create an image of the known word to memorise the association [99]. This association based technique provides a powerful tool for words that have a high degree of "imagenability" [105], or for word pairs between which the learner can form some kind of semantic link [42]. The important thing is that the keyword should clearly relate to the thing being remembered.

A wide range of existing studies in the broader literature have explored the effectiveness of the keyword method [14, 102, 11, 136]. For example, Atkinson and Raugh [14] found that participants who were given a keyword along with the translation learned more words and also remembered more words after 6 weeks. Also, Sagarra and Alba [107] compared rehearsal, semantic mapping displays and the keyword method, and found that the keyword method resulted in the best retention. It has also been shown that the keyword method is superior over systematic teaching [66, 100]. In the same sense, Raug et al. [103] evaluated the use of the keyword method over a long period of 8 to 10 weeks to teach Russian vocabulary and found it to be highly effective. Despite the generally positive results [99, 107, 66, 103] , there have been some negative findings [130, 136]. For example, a study conducted by Zheng Wei [136] found no significant differences between the keyword method, the word-part technique (recognizing part of a word) and the self-strategy. One specific concern is that the keyword method may be less effective when the target materials are not "keyword friendly"—that is, when they lack an obvious keyword or are difficult to visualise [50]. Until future research clarifies which factors influence the success of the keyword method (e.g., how much experience does the user have with learning languages), its use should be limited to keyword friendly materials.

Timing of Guidance

Different educational practices and learning perspectives have been also debating on "when" exactly the instructional support needs to be provided in a learning environment. While not always in agreement about when specifically the guidance should be given, all these disputes admit that the timing of instructional guidance is important [9, 115].

From one perspective, the best time to provide guidance is as soon as possible – i.e., "either at the beginning of the instruction or as soon as a learner makes an error". However, the detailed research on "intelligent-tutoring systems" suggests that depending on the instructional goals being pursued, providing immediate guidance is not always the best strategy [9, 85]. Here, "intelligent-tutoring systems" are computer-based problem-solving environments that model problems for learners and offer personalised guidance based on observing every step of their attempts at solving a problem [131].

A pioneering work by Anderson notes that offering learners guidance as soon as they deviate from a path to a viable solution, increases their problem-solving speed [9]. To test this hypothesis, the authors conducted an experiment with two conditions using an intelligent tutor – a computer systems that aim to provide immediate and customised or feedback to learners, without a human instructor [48]–, to teach students LISP programming [18]. In one condition, learners were interrupted during their work and offered guidance as soon as they took a step that would not lead to a correct solution. In the other, learners received guidance only on their request. Findings from this study indicate that learners who were interrupted during their work and offered guidance as soon as they deviate from a path to a viable solution completed the programming exercises in about half the time taken by those who received feedback only on request [9].

On the other hand, in a review of the intelligent-tutoring literature, Mathan

> and Koedinger suggest that delaying guidance may result in better retention and transfer of learning [85]. To test this hypothesis, Mathan and Koedinger also experimented using two different conditions. In one condition, the tutor offered immediate guidance on errors. In the other, the tutor waited to see whether learners detected their own errors, and attempted to guide them through detecting and correcting their mistakes only if they attempted to move on to a new problem. Findings from this study show that while learners in the two groups performed similarly on the first problem, those in the delayed guidance condition learned at a faster rate on all subsequent problems [85].
>
> These studies suggest two things about the timing of instructional guidance. First, if the guidance is constantly provided to the learner, timing strongly affects short-term outcomes. Second, the timing of guidance should vary according to instructional goals. While immediate guidance promotes more rapid problem-solving in the short term, delaying guidance can result in better long-term retention and transfer.

The literature reports that appropriate guidance helps increase user engagement in a learning activity [22, 128] and that engagement is important to support experiential learning [91]. Therefore, a better understanding of engagement in learning is needed when designing instructions for ExL environments. In the following section, we therefore provide a brief description of engagement and its dimensions (i.e., affective, cognitive, behavioural and agentic) that are useful for designing effective learning environments.

2.3 Engagement in Learning

Engagement is a necessary first step in learning [30]. Engagement is considered from different aspects in the literature since a common approach or a theoretical structure is lacking in relation to users engagement in educational environments [34, 45, 74]. However, the most comprehensive view emphasises that engagement is a complex and multi-dimensional process [112, 108, 128] intertwined with learners' internal indicators such as motivation, feelings, etc. (affective dimension) [46], mental effort, perceptions, etc. (cognitive dimension) [46], and observable actions such as performing various activities, interacting with a system, etc. (behavioural dimension) [12, 46]. In addition, Reeve and Tseng suggested incorporating agentic engagement as a fourth dimension of engagement [104]. Agentic engagement refers to the proactive and intentional activity of the learner to personalise the conditions of learning and to enrich external learning goals.

Previous studies have shown that behavioural, cognitive and affective dimensions predict learner' performance as separate dimensions and also together form a bigger construct [7, 128]. This means that the dimensions of engagement are interrelated and simultaneously affect human behaviour [46]. Altogether, engagement in learning experiences; helps learners increase their satisfaction, enhances motivation to learn, reduces the sense of isolation, and improves performance during the learning process [30, 22, 128]. In terms of technology, immersive environments such as Extended Reality (XR) can be identified as creative tools for designing engagement and simulating such learning experiences [128, 49].

2.4 Extended Reality (XR) Technology

Milgram and Kishino first introduced the virtuality continuum or reality-virtuality continuum concept in 1994 [89]. The continuum represents the full spectrum of technological possibilities between the entirely physical world or real environment and the fully digital world or virtual environment. The different sections of the continuum define how many real elements vs. digital elements are displayed, starting from the left end –the real environment– where 100% of what is displayed are real or physical objects and 0% are digital elements versus the right end –the virtual environment– where 100% of the objects displayed are digital and 0% are physical objects. It includes all current technologies that alter reality with computer-generated graphics as well as those yet to be developed. In a continuum, adjacent parts are almost indistinguishable, but the extremes are very different. Therefore, the exact limits of the various terms are debatable.

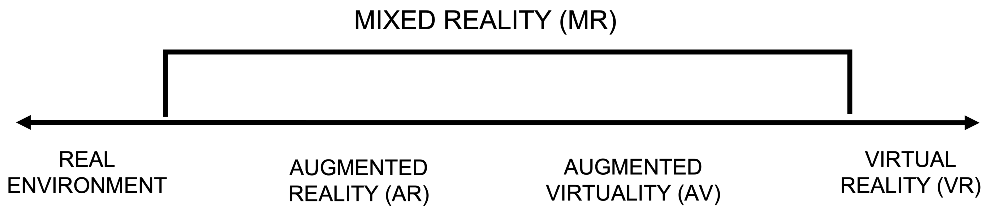


Figure 2.2: Reality-Virtuality continuum by Milgram and Kishino [89].

The virtuality continuum is broken down into four categories:

- Real environment: consists solely of real or physical objects. The real environment represents the left end of the virtuality continuum.
- Augmented reality: the real world is augmented with digital elements.
- Augmented virtuality: the virtual world is augmented by the inclusion of real or physical objects.
- Virtual environment: consists solely of digital objects. The virtual environment represents the right end of the virtuality continuum.

Extended Reality (XR) includes Augmented Reality (AR), Mixed Reality (MR), Virtual Reality (VR) and any other technology – "even those that have yet to be developed–situated at any point of the virtuality continuum"(Figure 2.3). AR has been described as "the concept of digitally superimposing virtual objects onto physical objects in real space so that individuals can interact with both at the same time" [15] (Figure 2.3 (b)). It enhances the real world with digital contents such as images, text, and animation. Individuals can access the environment through HMDs (AR glasses) or handheld devices such as tablets or smartphones. In contrast, VR is "an immersive, completely artificial computer-simulated environment with real-time interaction" [64] (Figure 2.3 (a)). An individual can experience and interact with the environment through a 360 view HMD. Finally, MR, is the result of blending the physical world with the digital world. In MR, digital and real-world objects co-exist and can interact with one another in real-time (Figure 2.3 (c)). It provides the ability to have one foot in the real world, and the other in an imaginary place, breaking down basic concepts between real and imaginary.

Extended reality has been utilised across multiple domains such as education [128, 124], psychology [133], construction [4, 122], medicine [41, 55] and the military [95, 80]. The main advantage of XR is that it can recreate environments that are difficult to simulate and provide a safer environment with less risk [95]. Other advantages of XR include the ability to control and manipulate constraints in complex and dynamic environments to create specific situations that are repeatable (e.g. landing an aircraft after an engine failure during a flight simulation). Compared to the traditional classroom learning method, XR has many advantages in the field of education. Instead of traditional whiteboard instruction, XR can create an immersive learning environment where learners can thrive. With an immersive environment, education becomes an experience that is fun and engaging. Learning in an immersive XR environment has also been proven to enhance the performance of the learners [132, 111, 128]. The interactive nature of XR brings elements to life, thereby creating a highly personalised user experience. With XR, learners can pay more attention to their surroundings and make connections to their experience, which leads to better knowledge retention.

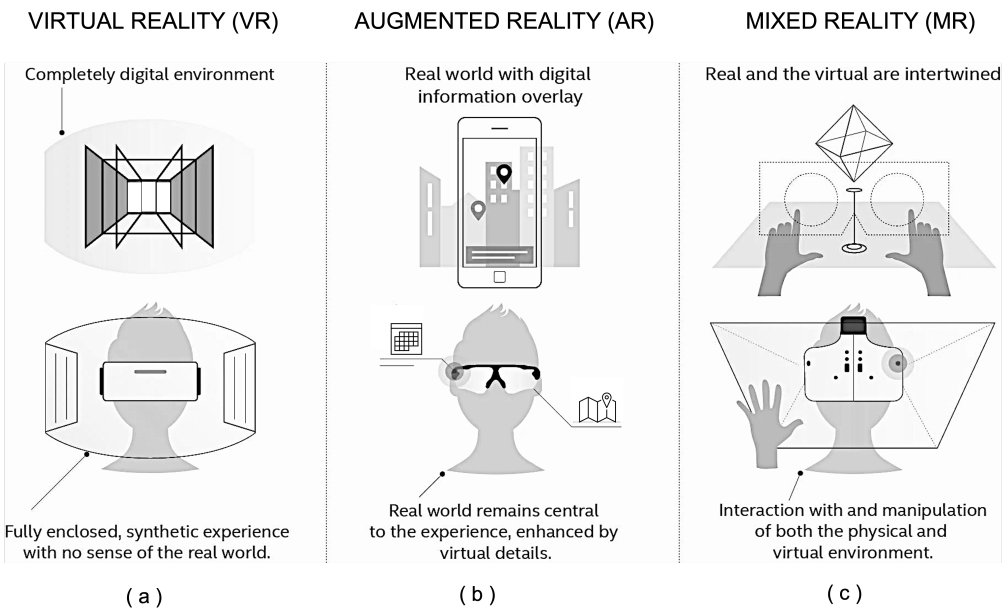


Figure 2.3: Extended reality environments: (a) Virtual reality, (b) Augmented reality, (c) Mixed reality.

2.4.1 Learning in XR

Researchers in the HCI community have leveraged XR systems as tools to guide or teach different skills and activities that can be learned best through experience [121, 58, 56, 57, 51, 128, 124, 17, 60, 8, 76]. However, these works have primarily focused on learning stages of the Kolb's ExL cycle. According to Kolb, the learner must continue cycling through all the four stages of the ExL cycle, thus creating a "learning spiral of ever-increasing complexity" in order to gain better learning outcomes [70]. Currently, we are unaware of any literature that focuses on providing guidance in XR to support learners move from one stage to another in the ExL cycle and progress on the learning spiral.

Furthermore, most of these work have provided fixed guidance, in which all the users receive the same set of general instructional cues, with the same amount, in the same way [121, 58, 56, 57, 128, 17, 60, 8, 76]. However, the theories and other groundworks on instructional guidance claim that the type, the amount and timing play a vital role when providing instructional guidance as they differently affect engagement and consequently the learning outcomes of each individual learner [9, 115, 139]. Currently, besides the limited numbers of studies such as [51, 124], there is a lack of research on providing adaptive guidance by adjusting the type, the amount and timing of the instructions for different user needs.

2.5 Chapter Summary

This chapter has described the background and research relevant to this thesis. In Section 2.1, we looked at the theories and model that describe the learning process, with a particular focus on experiential learning models. Section 2.2 has introduced the concept of instructional guidance and its broader position in designing adaptive instruction for experiential learning systems. Section 2.3 has provided a brief description of engagement and its dimensions that are useful for designing effective learning environments. Before concluding this chapter, Section 2.4 has outlined the novel technologies, such as AR, VR and MR, for developing immersive experiential learning systems and their existing literature.

As highlighted in Section 2.4.1, despite the fact that researchers have used AR, VR and MR technologies to develop learning systems, there are still some potentially open questions that need to be addressed, such as: How to design XR-based future experiential learning environments to enhance user engagement and learning performances? How to improve instructional guidance through "adaptation/personalisation"– i.e., remodel the type, timing and the amount, of guidance? Is it possible to develop adaptive XR learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators ? and, How the adaptation of guidance would affect the engagement and consequently the learning outcomes?. Therefore, to shed further light on this topic, the following chapter presents a broader design space with possible dimensions, developed based on the features of XR environments and the theoretical underpinnings of learning and guidance described in this Chapter 2.

Chapter 3

Design XR Learning Systems

> "Any sufficiently advanced technology is indistinguishable from magic.
>
> *Arthur C. Clark*

If novel technologies, such as Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), are to be more widely adopted as learning tools in the future, it will be necessary to consider how existing learning environments can be better supported. Therefore, the design of future learning systems needs to be considered from many perspectives, such as existing learning practices, established learning models that influence learning approaches, the different guidance methods, and the features that novel technologies can offer. Thus, in this chapter, we introduce a design space with several dimensions that can potentially be used to design and develop interactive learning systems which can be integrated into future classrooms and/or our day-to-day lives. Section 3.1 of this chapter presents the design space with its dimensions: –the display configuration, data modality, input, perceived space, experiential learning, and guidance – in more detail. Section 3.2 describes how the dimensions of our design space are used to systematically explore the existing body of literature and identify the gaps in knowledge. Finally, Section 3.3 describes how we utilised the presented design space to develop three different interactive learning systems to answer the research questions presented in Section 1.2.

3.1 Design Space

While previous work has included specific configurations of XR-based instruction to support experiential learning, this thesis aims to present a broader design space by exploring the domain, and highlighting the importance and effectiveness of different configurations based on the features of XR environments and the theoretical underpinnings of learning and guidance.

The performance of an experiential learning task takes place in a particular environment and follows a specific set of instructions. In addition to these fundamental dimensions, in our design space we explore how AR and VR can be used to augment and facilitate experiential learning and lead learners to progress in the experiential learning spiral. Therefore, in our work, we introduce the following six dimensions: *Display Configuration*, *Data Modality*, *Input*, *Perceived Space*, *Experiential Learning* and *Guidance* to form a design space that can be used to categorise previous work and/or to develop future learning systems.

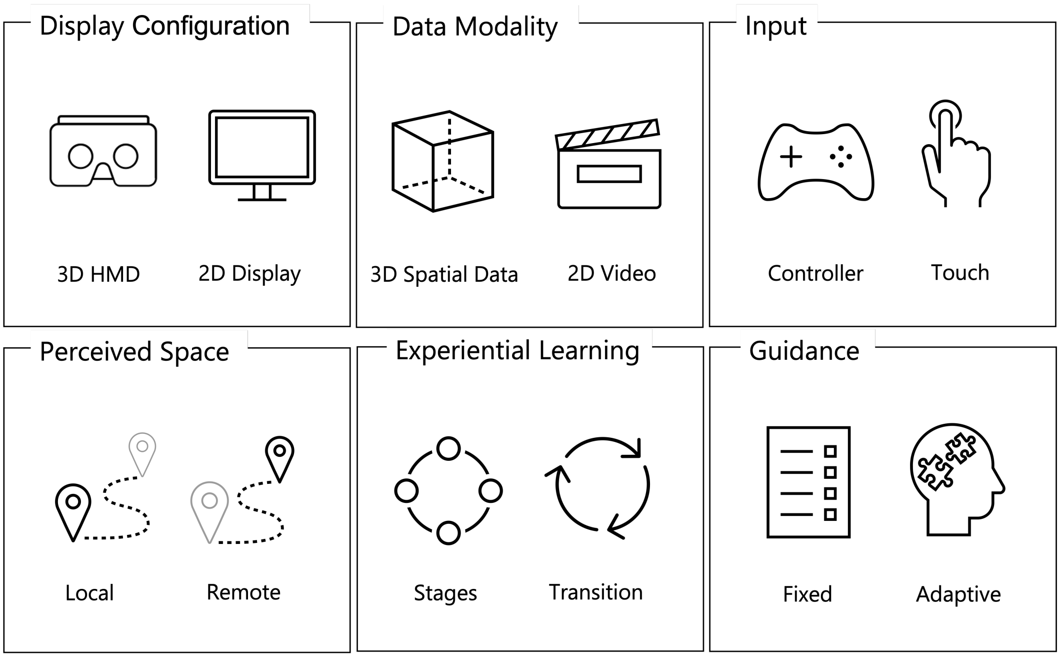


Figure 3.1: The dimensions of the design space.

Display Configuration

The *display configuration* dimension refers to "how" the user can see and interact with space(s), and environments, which can take many forms depending on the technology available. In this thesis, we explore *mixed reality (MR)* and *virtual reality (VR)* as a means to observe and interact with the user's perceived space. For instance, MR gives the user a direct view of the real environment, the ability to interact with it naturally, as well as the ability to augment and annotate the local space. On the other hand, VR can simulate both physically existing (natural or already built) and non-existing (imaginary or to be built) environments and allows the user to observe and interact with it immersively. Additionally, both MR and VR enables the user to situate a remote user in their own space as an avatar and interact with them as if they are actually present there. In our approach, we only focus on head-mounted displays (HMD), which enable MR (e.g. Magic Leap one [1] or Microsoft HoloLens 2 [2]) and VR (e.g. HTC Vive [3]) individually. Other display configurations, such as projected or hand-held video-see-through AR, and desktop VR are also possible, but outside the scope of our current investigation.

Data Modality

[1] https://www.magicleap.com/

[2] https://www.microsoft.com/hololens

[3] https://www.tobiipro.com/product-listing/vr-integration/

The *data modality* dimension refers to “what” type of data is obtains and is presents to convey the information in the learning environment. Using spatial data, the user could see and interact with the local or remote space using a 3D reconstruction of the environment, that is not readily available through other forms such as video. It also allows navigating to novel viewpoints, to avoid issues with occlusion, to add 3D annotations (e.g. 3D objects/animations) in the space, etc. Additionally, the user could also see a video of the space. The video can provide a high resolution, easily understood mechanism to comprehend the environment and activities. In the local space, multiple videos can provide in different viewpoints to enable third-person views of the user’s own actions and environment. Many other technologies can be leveraged to provide novel lenses to view and interact with the spaces (e.g. embedded sensors, recording audio), however, we limit our exploration in this thesis to the ones described above.

Input

The *input* dimension refers to “how the user can interract” with the space(s), and environments, which could also take many forms depending on the technology. XR platforms can take advantage of a rich variety of input features when designing user interactions. Mainly, these interactions can be designed in two different ways: as controller inputs or hand gesture/touch inputs. However, access to these input features can vary a lot between platforms. Other inputs, such as keyboard, voice, and eye gaze are also possible, but outside the scope of our current investigation.

Perceived Space

The *perceived space* dimension refers to “which” space(s), and environments the user can see and interact with. Each user has the potential to see and interact with their own *local space*, which is the environment that they are physically within and the objects within that space. The user would primarily interact in this space to perform the task or action in their own environment with their own objects or tools. The user may also see the other user’s *remote space*, which is the environment and objects of another user. In this space, the user can observe, inspect and give feedback on the remote user’s actions and their interactions with objects and tools.

Experiential Learning

The dimension of *experiential learning* refers to “which elements” of the ExL cycle we focus on. Kolb’s experiential cycle is formed of four *stages/bases* and four *edges* to progress between the stages. Designed instructional support in XR environments could be used to support these bases of the experiential learning cycle. Instead, it could also be used to encourage users to move from one base to the next. In this case, users can touch all bases, go through the experiential learning cycle and progress towards the learning spiral.

Guidance

The *guidance* dimension refers to “what type of instructional support” is created to assist the learners to move from one stage to another in the experiential learning cycle. The type of instructional support could be *fixed*, in which all the users get the same set of generic instructional cues in the same amount, at a specific predefined time slot. At the same time, it could be a *adaptive* set of instructions based on the user needs and the user engagement (cognitive, affective and behavioural engagement). In this case, the type (generic to mnemonic techniques) and the amount (minimally guided to fully guided) of the instructional is support remodelled according to the learning performances and cognitive capabilities, such as “cognitive load” and attention, of different users.

Here, “cognitive load” relates to the amount of information that working memory can hold at one time [120]. However, *Cognitive Load Theory* shows that working memory can be extended in two ways. First, as the mind processes visual and auditory information separately, auditory items in working memory do not compete with visual items in the same way that two visual items compete with each other (e.g. a picture and some text compete with one another). Second, working memory treats an established schema as a single item. So, learning activities that draw upon prior knowledge expand the capacity of the working memory. Therefore, pre-training, or teaching learners prerequisite skills before introducing a more complex topic, will help them establish schemas that extend their working memory; which then leads them to understand and learn more difficult information easily.

3.2 Learning in XR

	Display Configuration	Data Modality	Input	Perceived Space	Experiential Learning	Guidance
Botelho [19] Menaker [87] Adams [3]	Desktop	2D Video	Keyboard	Local	Stages	Fixed
Ternier [121] Huang [58] Hsu [56]	Handheld	3D Data	Touch	Local	Stages	Fixed
Huang [57]	AR Glass	3D Data		Local	Stages	Fixed
Herbert [51]	AR	3D Data	Controller	Local	Stages	Adaptive
Alrehaili [8] Kwon [76] Topu [128]	VR	3D Data	Controller	Local	Stages	Fixed
Jarmon [60]	VR	3D Data	Controller	Local + Remote	Stages	Fixed
Birt [17]	AR + VR	2D + 3D	Controller	Local	Stages	Fixed
Thoravi [124]	AR + VR	2D + 3D	Controller	Local + Remote	Stages	Adaptive

Table 3.1: Prior work on learning in XR as they fall within the design space illustrated in Figure 3.1.

In this thesis, we first use above mentioned dimensions in our design space (as illustrated in Figure 3.1) to categorised the prior work on learning in XR that presented in Section 2.4.1. It is summed in Table 3.1. Notably, most of the previous work makes a single set of choices along these dimensions and often centred on offering fixed guidance, in which all the users receive the same set of general instructional cues, in the same way, with the same amount, at a specific predefined time frame [128, 93, 17, 60, 119, 76]. According to the debates about the impact of instructional guidance presented in Section 2.2, providing appropriate information in the right amount, at the right time, and in the right format is important, as it affects engagement and the learning outcomes of each individual learner differently. However, besides a few recent AR, VR studies such as [51, 124], there is a lack of research on offering adaptive guidance by adjusting the type, the amount and timing of the instructions for different user needs.

Further, these works have primarily focused on creating simulated environments to support experiential learning tasks [121, 56, 58, 57, 51, 8, 76, 128, 17, 60, 124]. A key insight from our background review of the learning literature (Section 2.1) is that experiential learning tasks involves several distinct stages – doing or having a novel experience; reviewing and reflecting on the novel experience, focusing on whether there were any discrepancies between the experience and understanding; what was learned from this experience, thinking how it can be improved for the next time; and planning or practically applying the learning. As described in Section 2.1 Kolbs' experiential learning model, within a single learning session, the ability to switch spontaneously between these stages and continue cycling through all the stages of the experiential learning cycle is beneficial for obtaining a better learning outcome [70]. However, in previous works, we could not find any AR, VR or MR study that focuses on supporting learners to move spontaneously from one stage to another in the experiential learning cycle and progress on the learning spiral.

3.3 Our Learning Systems

Our general problem in this thesis is how to use XR technologies to design and develop adaptive learning systems to improve learning and promote engagement through active, stimulating, and interactive environments supported by appropriate guidance. After reviewing the existing body of literature and identifying the gaps, we used the dimensions of our design space (in Figure 3.1) to design and develop three different interactive learning systems viz. VRNav, Arigatō and VocabulARy, that differ in particular on *display configuration, experiential learning and guidance* dimensions of the design space (Figure 3.2). Further, the three systems also have different learning objectives. For example, VRNav focuses on spatial learning in which the learners acquire a mental representation of the environment by experiencing it through various representations. Spatial learning is essential for individuals to navigate and interact effectively with their surroundings [24]. On the other hand Arigatō and vocabulARy focus on second language learning in particular on learning vocabulary and grammar. Despite the different learning objectives across these three studies, they all share the common ground on the potential benefit from experiential learning. Experiential learning emphasises the active engagement of learners in practical experiences, enabling them to construct their knowledge through reflection and application [69]. For example, a study conducted by Kozhevnikov et al. [73] investigated the impact of experiential learning on spatial ability. They found that engaging in immersive virtual environments that require spatial navigation and exploration significantly improved participants' spatial skills and mental representation abilities. A meta-analysis by Elola and Oskoz [43] examined the effectiveness of different instructional approaches for second language vocabulary learning. They found that experiential learning methods, such as using multimedia, authentic materials, and real world contexts, were effective in promoting vocabulary acquisition and retention. As such it is important to know how one can simulate and enrich such experiential learning environments with novel technology such as XR.

In order to evaluate the above mentioned learning system we designed specific user studies that considered qualitative and quantitative data collection techniques, which would allow us test the effects of different configurations of XR and degree of dynamic adaptions of guidance on learning outcomes and user engagement and answer the research questions in Section 1.2.

	Display Configuration	Data Modality	Input	Perceived Space	Experiential Learning	Guidance
VRNav (Publication 01)	VR + Desktop	2D + 3D	Controller + Keyboard	Local	Stages	Fixed
Arigatō (Publication 02)	AR	2D + 3D	Controller	Local	Stages + Transition	Adaptive
VocabulARy (Publication 03)	AR + Tablet	2D + 3D	Gestures + Touch	Local	Stages	Adaptive

Table 3.2: (i) VRNav, (ii) Arigatō and (iii) VocabulARy learning systems as they fall within the design space illustrated in Figure 3.1.

VRNav (Publication 01)

VRNav presents two prototype systems developed for a VR HMD (i.e. Tobii Pro HTC Vive [4]) and a regular computer display (i.e. Tobii pro spectrum computer display [5]). Each prototype grants a different representation method for the same 3D virtual tour of a building space. In both systems, the experiential learning task was to freely explore the environment without any additional guidance. However, to facilitate navigation, information boards and signs were included in the environment as in an actual building. The two virtual environments differed only in terms of *Display configuration* (VR HMD and Desktop), and *Input* (HTC controller and computer mouse) dimensions (Figure 3.2).

This study is centred around our first research question, *RQ1:How can XR technology be used to create immersive and interactive experiential learning environments by exploring the importance and effectiveness of different configurations of XR, ExL and instructional guidance?*, and its corresponding hypotheses presented in the Section 1.2. In the *VRNav* study, focus is placed on how VR HMD

[4] https://www.tobiipro.com/product-listing/vr-integration/
[5] https://www.tobiipro.com/product-listing/tobii-pro-spectrum/

and the other representation methods (desktop VR and paper-based 2D floor plans) of the office building would (i) affect the dimensions of user engagement and the understanding of the space with minimum guidance (i.e. the spatial and the volumetric understanding of the building) and, (ii) support the users project themselves into the building space.

Arigatō (Publication 02)

Arigatō presents a prototype system for language learning developed on the Magic Leap one [6] AR HMD (Figure 3.2 (c)) with controller inputs. The *Experiential Learning* (Stages and Transitions) and *Guidance* (Adaptive Amount and Fixed Amount) dimensions were mainly considered when designing the *Arigatō* prototype (Figure 3.2).

The prototype replicates all four stages of Kolbs' experiential cycle. The system starts the learning cycle in the abstract conceptualisation (AC) stage. The AC stage provides the description of the vocabulary and the basic grammar rules related to phrases being learnt. The vocabulary is presented using two different types of instruction methods (i.e, (i) fixed-associations– 3D virtual objects of real world objects associate with the vocabulary or (ii) adaptive-associations– merged 3D virtual objects/combined metaphors created using the keyword method). After familiarising themselves with the content in the AC stage, the learner proceeds to the active experimentation (AE) stage.

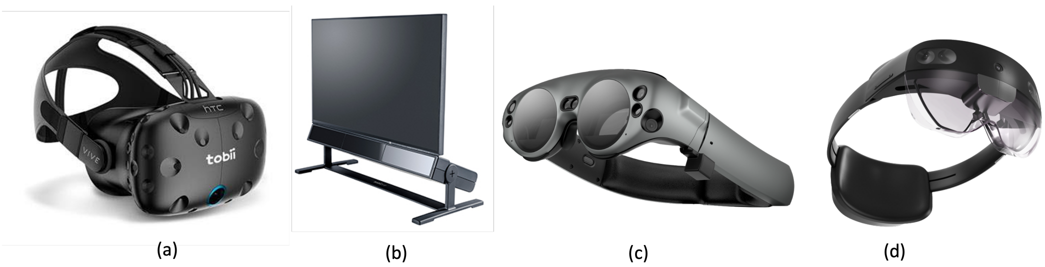


Figure 3.2: Hardware used: (a) Tobii pro HTC vive HMD, (b) Tobii pro spectrum display, (c) Magic leap one HMD, (d) Microsoft hololens 2 HMD.

In the AE stage, the system shows each English phrase with a puzzle task to generate the corresponding Japanese phrase made of three parts from six (6) possibilities offered. The puzzle is presented in two ways based on the type of instructions: (i) a word puzzle with text components only or (ii) a word puzzle together with 3D virtual objects for the learner to create associations using the keyword method from a given set of objects. If the learner does not know the solution to the puzzle they can skip that phrase.

After completing the puzzles related to all four phrases, the system nudges the learner to proceed to the concrete experience stage (CE). A 3D avatar is shown asking the learner to recall each phrase in Japanese by speaking it aloud. To evaluate the spoken answers, the IBM Watson speech to text recognition was used. If the phrase is correctly voiced, the corresponding 3D AR model is displayed. The learner can manipulate the object and place it anywhere in the room as they

[6] https://www.magicleap.com/

would be able to do it in a real-world experiential learning scenario with physical objects.

If all the phrases are not recalled correctly in the CE stage, the system automatically moves to the reflective observation (RO) stage. There, the learner can first go through all the phrases that include corresponding visual models and audio pronunciation, and then recall them by selecting the correct Japanese phrase (out of four phrases) corresponding to the given English phrase. The system then returns back to the AC stage. Besides text instructions and explanations, audio instructions and explanations are also provided throughout the cycle. In this way, the system nudges the learner to proceed further in the experiential learning cycle and consequently on the learning spiral until all four (4) phrases are recalled correctly in the CE stage.

Arigatō explores *RQ2: What are the XR interventions that can be used to remodel the type and the amount of instructional guidance?, RQ3:How can XR technology be used to create immersive and adaptive experiential learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators?, and RQ4:How do the different configurations of XR and adaptions of guidance would affect the engagement and learning outcomes of immersive ExL environments?*, presented in Section 1.2. In particular, the aim of this study is to: (i) identify the XR interventions that can be used to "personalise" – i.e. remodel the type and the amount of – instructional guidance and (ii) investigate how the different adaptation factors of instructional guidance, – i.e. the amount (fixed vs. adaptive-amount) and the type (fixed vs. adaptive-associations) of guidance – would affect user engagement and, consequently, the learning outcomes of a language learning scenario.

VocabulARy (Publication 03)

VocabulARy presents two prototype systems for vocabulary learning developed on an AR HMD (i.e. Microsoft HoloLens 2 [7]) and an 10.5 *in* Android tablet device (i.e. Samsung Galaxy Tab S4 [8]). Both AR and tablet systems combine the *keyword method* (Section 2.1) with physical objects in the context. With the AR HMD, the system allows the user to interact with the real-environment where certain objects are labelled with a button indicating that their translations are available. Upon clicking these buttons with a hand gesture, the English and Foreign words corresponding to the object appear. At the same time, an audio of the pronunciation of the foreign word is also played. In addition to the words and audio, a pre-generated keyword with an animated 3D visualisation, also appear. The functionalities of the tablet prototype is similar to the AR prototype. However, instead of seeing the real world environment, an image of an environment is displayed on the screen. The *Display Configuration* (Tablet and AR HMD) and *Guidance* (Keyword and Keyword+Visualisation) were the main dimensions considered for designing the *VocabulARy* system (Figure 3.2).

VocabulARy investigates RQ3 and RQ4 (as discussed above), presented in Section 1.2. More specifically, the aim of study is to investigate how (i) different

[7] https://www.microsoft.com/hololens
[8] https://www.samsung.com/si/tablets/galaxy-tab-s/

display configurations – i.e. interfaces (AR vs. non-AR) – and (ii) different types of guidance – i.e. instruction mode (keyword only vs. keyword + visualisation) – would affect user engagement and, consequently, the learning performances in vocabulary learning.

3.4 Chapter Summary

This chapter has presented a comprehensive understanding of the dimensions of a design space and showed, how it was (i) used to systematically explore the existing body of literature and identify knowledge gaps, and (ii) utilised to develop three different interactive learning systems (*VRNav*, *Arigatō*, 6) to answer the research questions presented in Section 1.2. The next three chapters (Chapter 4,Chapter 5,Chapter 6) include the research articles that we published based on the studies described in Section 3.3. In particularly, Chapter 4 includes the journal article published based on the Study 1: *VRNav*, that was conducted to investigate how different representation methods (VR, paper and desktop) would affect the dimensions of user engagement and the understanding of the space with minimum guidance. Chapter 5 presents the article published based on the Study 2: *Arigatō*, that explores the effects of adaptive guidance on engagement and performance in AR learning environments. Chapter 6 includes the journal article published based on the Study 3: *vocabulARy*, that investigates the effects of different types of guidance on engagement and performance of vocabulary learning in AR.

Chapter 4

Publication 01: VRNav [135]

Title: Exploring the Future Building: Representational Effects on Projecting Oneself into the Future Office Space

Authors: Maheshya Weerasinghe, Klen Čopič Pucihar, Julie Ducasse, Aaron Quigley, Alice Toniolo, Angela Miguel, Nicko Caluya, Matjaž Kljun

Year: 2022

Journal: Virtual Reality (Q1)

DOI: 10.1007/s10055-022-00673-z

Link: https://link.springer.com/article/10.1007/s10055-022-00673-z

Summary: The VRNav Study has two aims. The first aim is to investigate the impact of immersion by projecting participants into a future scenario, while the second aim focuses on the learning process (in particular spatial learning). The analysis of immersiveness in the first part is not intended to measure learning directly, but rather to serve as an indicator of the persuasive and effective nature of immersive environments in simulating experiential learning. While the objective of the second part is to enable participants to develop a mental representation of the environment through spatial learning (by experiencing the environment), facilitating various tasks such as distance, direction, and volumetric estimation. To this end, we explored the impact of different *Display Configurations* and *Data Modalities* on user engagement and understanding of space with minimal guidance. We observed and evaluated different representation methods –i.e., VR, desktop and 2D paper-based floor plans – on spatial comprehension in ExL environments, through various data collection and standard questionnaires combining qualitative and quantitative research methods. Through this the study directly addresses research question RQ1 (How can XR technology be used to create immersive and interactive experiential learning environments by exploring the importance and effectiveness of different configurations of XR, ExL and instructional guidance?). The findings from this study provided valuable insights into the impact of immersive VR and other representation methods on user experience, user engagement dimensions, and spatial understanding of ExL environments. Additionally, these findings helped to address the research question RQ4 (How do the different configurations of XR and/or degree of dynamic adaptations of guidance affect the user experience, engagement and, consequently, the learning outcomes of immersive ExL environments?).

Virtual Reality
https://doi.org/10.1007/s10055-022-00673-z

ORIGINAL ARTICLE

Exploring the future building: representational effects on projecting oneself into the future office space

Maheshya Weerasinghe[1,2] · Klen Čopič Pucihar[1,3] · Julie Ducasse[1] · Aaron Quigley[4] · Alice Toniolo[2] · Angela Miguel[2] · Nicko Caluya[5] · Matjaž Kljun[1,3]

Received: 13 April 2021 / Accepted: 21 June 2022

Abstract
While virtual reality (VR) has been explored in the field of architecture, its implications on people who experience their future office space in such a way has not been extensively studied. In this explorative study, we are interested in how VR and other representation methods support users in projecting themselves into their future office space and how this might influence their willingness to relocate. In order to compare VR with other representations, we used (i) standard paper based floor plans and renders of the future building (as used by architects to present their creations to stakeholders), (ii) a highly-detailed virtual environment of the same building experienced on a computer monitor (desktop condition), and (iii) the same environment experienced on a head mounted display (VR condition). Participants were randomly assigned to conditions and were instructed to freely explore their representation method for up to 15 min without any restrictions or tasks given. The results show, that compared to other representation methods, VR significantly differed for the sense of presence, user experience and engagement, and that these measures are correlated for this condition only. In virtual environments, users were observed looking at the views through the windows, spent time on terraces between trees, explored the surroundings, and even "took a walk" to work. Nevertheless, the results show that representation method influences the exploration of the future building as users in VR spent significantly more time exploring the environment, and provided more positive comments about the building compared to users in either desktop or paper conditions. We show that VR representation used in our explorative study increased users' capability to imagine future scenarios involving their future office spaces, better supported them in projecting themselves into these spaces, and positively affected their attitude towards relocating.

Keywords Immersive VR environments · User engagement · Sense of presence · User experience · Job relocation

1 Introduction

The collaboration between people from the fields of architectural design, building construction, and various other stakeholders such as investors and future occupants is pivotal for the successful completion of construction projects. However, communication failures and misunderstandings can easily compromise such projects. In recent decades, computational technologies such as building information modelling (BIM) have been used to assist the construction industry (Azhar 2011; NBS 2016). BIM is defined as the "use of a shared digital representation of a built asset to facilitate design, construction and operation processes to form a reliable basis for decisions" (ISO 2018).

Although not strictly necessary (Leon 2014), a BIM process can run throughout the entire life cycle of a construction project from the early conceptual design phase well into the time when the building structure is already in use. The conceptual design phase (activities of function formulation, concept generation, concept organisation, concept evaluation, and concept improvement) is a very important stage in the entire process in order to bring everyone on board to

✉ Maheshya Weerasinghe
amw31@st-andrews.ac.uk

1 University of Primorska, FAMNIT, Koper, Slovenia

2 School of Computer Science, University of St Andrews, St Andrews, UK

3 Faculty of Information Studies, Novo Mesto, Slovenia

4 School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

5 Interactive Media Design Laboratory, Nara Institute of Science and Technology, Nara, Japan

understand all the future implications of the finished project (Meng et al. 2020). However, when it comes to future office buildings, the number of studies involving future occupants in the conceptual design phase is scarce (some notable examples include (Westerdahl 2006; Frost and Peter 2000)). This is despite the fact that job relocation is one of the most stressful events people face and an employee's willingness to relocate depends on several factors including, amongst others, their perceptions and attitudes (Szpunar 2010).

Currently, the most commonly used formats of representation in the conceptual design phase are two-dimensional sketches and computer-aided design (CAD) models (NBS 2016). One of the biggest challenges of this phase is to clearly convey ideas in three dimensions (3D) through two-dimensional (2D) sketches and drawings. While professionals are used to mentally translating allocentric[1] to egocentric views (Kuliga and James 2020), it can be hard for someone outside the design world, to fully grasp the real implications of allocentric representations presented during pre-occupancy evaluation – i.e. before the building is constructed (Frost and Peter 2000). Therefore, the representation used in the conceptual design phase has a significant impact on and defines the efficiency of the information sharing, understanding of the architectural design, the overall quality of the design process, and the final outcome (Acock 1985). Virtual reality (VR) has consistently been shown to be efficient in pre-occupancy evaluation (Bassanino et al. 2010; Chandrasegaran 2013; Frost and Peter 2000; Westerdahl 2006; Kuliga and James 2020).

One of the significant features of VR environments is their ability to facilitate the sense of presence (Schwind et al. 2019; Benyon et al. 2014; Sanchez-Vives and Slater 2005; Witmer and Singer 1998) – the feeling of being or operating in a place by projecting oneself into that space while being physically present in another location (Banos 2004; Regenbrecht et al. 1998). VR thus supports users in gaining a critical perspective of 3D virtual models by experiencing the sensation of moving around the space, understanding its composition, viewing details from various angles (allo- and egocentric), and reacting accordingly, to shape the next steps of construction projects. Most commonly, moving is achieved either by walking or teleporting. Compared to the latter, walking delivers a more immersive experience (Slater et al. 1995; Usoh et al. 1999), reduces cognitive load (Zanbaka 2005) and VR sickness (Llorach et al. 2014; Jaeger and Mourant 2001). While teleporting is considered risk free (especially in confined physical spaces), and it reduces the possibility of VR sickness (Keshavarz et al. 2015), it lacks optical flow, which lowers the sense of presence (Bowman et al. 1997), and may cause spatial disorientation (Bowman et al. 1997; Bakker et al. 2003).

With current advancements in VR technologies, many architectural studios and construction companies have extended their conventional 2D technical drawings into 3D immersive environments (Dajana 2019). It has been shown, that the use of immersive VR environments in the design and construction industry facilitates better visual perception, and forms conditions for collaborative decision-making and problem-solving in the conceptual design phase (Dunston et al. 2011; Frost and Peter 2000). Nevertheless, there is little known about how such representation of the office space or the involvement of future occupants in the process, affects their capability to project themselves into the future office space and willingness to relocate. Thus, the research questions that led our study are:

- RQ1 How does the representation method influence the projection of participants into their future office space in aspects such as, their engagement (RQ1a), sense of presence (RQ1b), and user experience (RQ1c)?
- RQ2 How does the representation method influence the perception of and imagining oneself in the future office space?
- RQ3 Are there any correlations between the participants' engagement, sense of presence, and user experience of the functional space, and can observing them together better support the results obtained?

In order to answer these questions we teamed up with the InnoRenew Centre of Excellence (CoE) institute that was in the initial phase of the construction process of a new office building for researchers and administrators. We developed three different representation methods: (i) the conventional paper-based 2D floor plans and rendered images of the future building (paper condition), (ii) a virtual environment experienced on a computer monitor with the highly-detailed 3D representation of the future building (desktop condition), and (iii) the same environment experienced with a head-mounted display (HMD) (VR condition). In particular, this explorative study aimed to answer our research questions by analysing the patterns of exploration, user engagement, and by analysing questionnaires for the sense of presence and user experience.

The main contributions of this article are: (i) an explorative study of how future occupants project themselves into their future office space based on the representation method of the virtual building (paper, desktop, VR), (ii) how this influences their attitudes towards the building and moving, and (iii) a comprehensive analysis of variables including user engagement, sense of presence, user experience as well as correlations between them in the context of (i) and (ii).

[1] For the purpose of this work we use the term allocentric view as one which is a top-down view of an object, such as a map (Kuliga and James 2020; Cartillier et al. 2021; Dey et al. 2014).

The next section covers the research background on VR in architecture, the engagement and sense of presence within VR, and job relocation. Section 3 describes the research method with three different conditions (paper, desktop and VR), participants and the study procedure. The results section (5) is followed by discussion (4) and conclusion (6).

2 Research background

VR in architecture and construction provides the conditions for greater collaboration among stakeholders (Bassanino et al. 2010; Berg and Vance 2017; Fernando et al. 2013), allows more solid design decisions, and supports identification of the design and development issues that might not be identified efficiently in conventional ways (Bassanino et al. 2010; Dunston et al. 2011; Frost and Peter 2000). In addition, VR can provide designers and researchers with reliable user behaviour during human-building interactions (Bassanino et al. 2010; Kuliga and Thrash 2015; Heydarian et al. 2014), and can enhance safety training (Xie et al. 2006) by providing high levels of immersion that can optimise the learning process (Faas 2014). With the use of VR, both the cost and time spent on decision-making by building physical mock-ups used in the design review process can also be significantly reduced (Majumdar et al. 2006; Juan et al. 2018). Overall, VR provides users with a realistic perception of the design (Fernando et al. 2013) as well as the ability to "simulate the experience of moving through and interacting with the virtual world as if it was real" (Bassanino et al. 2010, p. 3).

Most of the studies of VR in architecture and construction focused on cognitive or affective aspects separately (e.g. (Berg and Vance 2017)), and used either simple abstract representations of space (non highly realistic) or environments based on semi-immersive projections (Schnabel and Kvan 2003; Westerdahl 2006; Ruddle et al. 1999). Related works to ours include a comparison between VR and a real building (Westerdahl 2006) and desktop and VR (Ruddle et al. 1999). The former study revealed that VR supports the decision-making process about the future workplace. The latter study showed that users in VR moved quicker and had a more accurate sense of relative straight-line distance. However, these works did not focus on how representation methods helped future occupants project themselves into the future office space.

We took the engagement, sense of presence, and user experience as the measures of projecting oneself into a future office space. We present these measures in the following subsections together with the related work on job relocation, stress and projecting oneself into the future.

2.1 User engagement

User engagement has been defined and studied in several contexts (Semiha 2019; Fredricks et al. 2011; Topu and Goktas 2019; Schaufeli 2013; Pierce et al. 2017). It is a dynamic, complex and multi-dimensional process (Schaufeli 2013; Pierce et al. 2017; Topu and Goktas 2019) composed of (intertwined) users' internal indicators such as feelings (affective dimension), thinking, reasoning, learning, etc. (cognitive dimension), and observable actions such as performing various activities, interacting with a system, navigating, exploring, etc. (behavioural dimension) (Appleton et al. 2008).

Kearsley and Shneiderman have stressed that engagement can be obtained without technology, but technology opens up novel possibilities that are hard to achieve in real life (Kearsley and Shneiderman 1998). Accordingly, research in this area has explored engagement with images (Frantzidis 2010), video clips (Murugappan et al. 2008; Yazdani et al. 2009), music (Takahashi and Akinori 2003; Koelstra 2012), and real-life scenarios (Katsis 2008; Weber 2019). With the advancement and availability of VR technologies, this medium has also been used and evaluated in a number of studies as a means of presenting emotional stimuli (Moghimi 2016; Violante et al. 2019; Birenboim 2019; Cebeci et al. 2019).

One of the foci of our first research question (RQ1a) is *How does the representation method of users' future office space influence participants' engagement?* Previous work has shown that when studied separately, affective, cognitive and behavioural dimensions affect user engagement in different contexts (Pierce et al. 2017), and that studied together they affect the engagement in an educational context (Topu and Goktas 2019). There is a lack of studies of these three dimensions studied together in the architectural design comparing different representation methods. We hypothesise that the method of representation will affect all dimensions of user engagement of future occupants while projecting themselves into their future office space.

2.2 Sense of presence and user experience

User Experience describes the subjective, holistic, situated sentiments of users towards a service, software or product in use (Marc 2010; ISO 2010). It thus includes a person's perception, emotion, cognition, motivation and behaviour – aspects that are inter-playing with the context of use (the place, time (before, during and after an interaction), people, and other objects) (idem). This understanding of user experience has expanded beyond the realm of human-computer interaction (HCI) and it is also used in architectural design (Krukar et al. 2016). HCI evaluation methods have been for example implemented to find functional deficiencies in buildings as products (Christoph 2006).

Recent user experience models divide user experience into two dimensions: pragmatic, that includes the objective properties of the service, software or product in use, and hedonic properties, that describe users' perception of these (Law et al. 2007). In the past decade both dimensions have been measured with questionnaires such as the popular UEQ (Schrepp et al. 2017; Schrepp 2019). It has been suggested that user experience in VR is affected also by the sense of presence (Steuer 1992) and the evidence supports a relation between the two (Busch et al. 2014; Brade 2017). The sense of presence can be described as the feeling of being present or immersed in a given setting upon the perception of it (Gifford 2007; Steuer 1992; Witmer and Singer 1998).

It's the immersion in VR enabled by stereoscopy, wide field of view, and a high degree of interactivity that facilitates a higher sense of presence (Dunston et al. 2011; Castronovo et al. 2013). Furthermore, since the human perception process is driven on inputs from different sensors (Gifford 2007; Steuer 1992), the more detailed the interface of the immersive environment, the more likely it is to enable a higher sense of presence (Bertol 1996). In an architectural setting, VR can be considered as the medium that enables users to explore the insides and outsides of a future building through its virtual 3D representation (Steuer 1992).

The importance of sense of presence as an outcome of immersive 3D VR environments has been highlighted in existing literature (Barfield and Weghorst 1993; Witmer and Singer 1998), and multiple questionnaires to measure it have been suggested (Witmer and Singer 1998; Mel et al. 1998; Usoh et al. 1999; Usoh and Catena 2000). The questionnaires used today have been carefully designed and refined over more than two decades. However, the use of questionnaires alone might not be enough to identify differences between the sense of presence in various representation methods (Mel et al. 1998; Slater and Martin 1993), as the sense of presence appears to be affected by a combination of environmental and personal factors (Witmer and Singer 1998).

Despite existing research on the user experience and sense of presence in 3D VR environments, there are still gaps in the knowledge (Renner et al. 2013). Several studies have focused on finding differences and similarities in (measuring) the sense of presence in VR and the real world (Westerdahl 2006). It has been confirmed that despite receiving different stimuli from VR and the physical surroundings, VR can still increase the sense of presence as users can filter information and focus by using selective attention (Witmer and Singer 1998; Pashler 1999). The comparison between playing a game on a flat display and a head-mounted display also demonstrated that there is a difference in experiencing the space and the perceived degrees of the sense of presence between the two (Federica et al. 2019). Nevertheless, immersive gaming presents a complex task-based activity, while the focus of this study is on a free-style exploration of the environment. Differences in sense of presence between desktop and VR have also been found in education (Makransky et al. 2019; Zhao et al. 2020).

The second and third foci of our first research question (RQ1b and RQ1c) is *How does the representation method of the model of their future office space influence the participants' sense of presence and user experience in it?* Based on previous studies we expect that the more immersive the environment, the better the user experience, and higher the degree of the sense of presence.

2.3 Job relocation

Job relocation has been mostly studied in the context of individual employees and is often associated with stress (Martin 2000), characterised by changes in physical environment, daily routines and social circles. The willingness to relocate depends on familial or background factors (single-earner marriage), personal factors or perceptions and attitudes (relocation and normative beliefs, self-efficacy, relocation policy satisfaction, organisational commitments and desire for career progress), spouse attitudes (willingness to relocate and relocation policy satisfaction), social factors (disruption of social contacts, social support), organisational factors (job characteristics) and others (Eby Lillian and Russell 2000; Luo and Cooper 1990; Maike and Margenfeld 2015). It is thus not surprising that relocation is often faced with resistance (Eby Lillian and Russell 2000).

When it comes to company relocation the studies are more scarce (Sagie et al. 2001). Relocation preparation (Robin 1999) as well as involvement of all employees (Chevi 2018) can simplify the relocation and reduce resistance. As mentioned, when a company is building new premises one of the possibilities is to involve employees by letting them participate from the design process onward and experience the building upfront. Involving users in the design process can help them project themselves into the future.

Projecting oneself into the future is a vibrant research area (see (Klein 2013) for a comprehensive review) also known by other terms such as future-oriented mental time travel or episodic future thought. It is the ability to imagine and simulate personal events that may potentially take place (Szpunar 2010). Recent studies have shown that past experiences are closely related to the ability of imagining one's future (Szpunar 2010; Schacter 2012). It is thus expected that exploring different representation methods with the same level of details but with different degrees of immersion will variously affect the experience and thus provide a different basis for imagining one's future. We hypothesise that the more

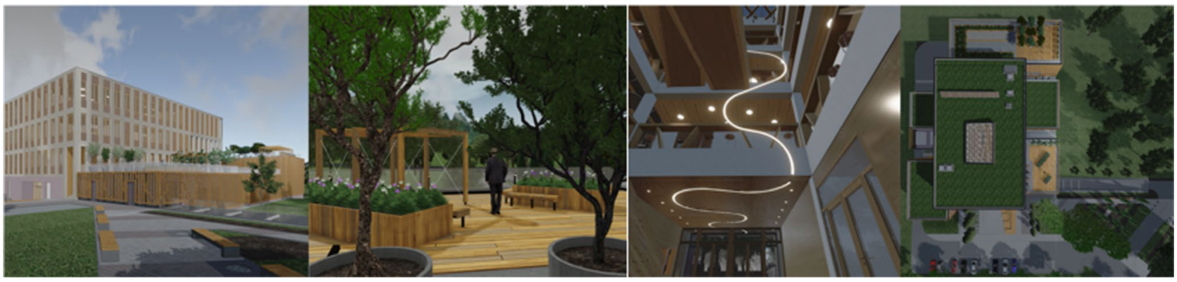


Fig. 1 Scenes from the Innorenew CoE building 3D environment

immersive the experience, the easier it will be for the future occupants to imagine how moving will affect them.

3 Research method

In order to investigate our research questions, we teamed up with architects and investors working on the future InnoRenew CoE[2] headquarters. The management of the institute agreed to participate in the development of the virtual experience and the explorative study, which happened in the final stages of the design process when high-fidelity mock-ups of the actual building were expected. This section describes the study conditions, study procedure, participants' sampling, data collection and how it was analysed.

3.1 Study conditions

We created three different conditions, each employing a different representation method of the future office space: (i) a paper based folder with 2D floor plans and rendered figures of the future office – paper condition, (ii) a 3D virtual experience presented on a regular computer monitor – desktop condition, and (iii) the same 3D virtual experience presented on the head-mounted display (HMD) – VR condition.

For the virtual experience, we used the 3D model of the building created by the architects using Autodesk 3ds Max, and developed it into a 3D virtual environment using the Unity 3D game engine. During the entire process we worked together with architects who provided their input on different aspects of the experience in order to make it as close as possible to how they envisioned it. Some of the scenes of the virtual experience are visible in Fig. 1. The building is four stories high (referred to as ground floor and floor 1 to 3) with floors connected by an internal staircase. Besides allowing the exploration of building interiors, the experience also allowed outdoor exploration up to 50 metres away from the entrance side of the building and up to 5 metres around other sides.

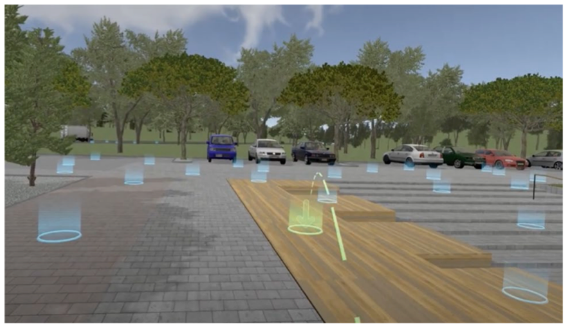


Fig. 2 Teleport points in the virtual environment

Desktop and VR conditions differed only in rendering type (non-stereoscopic vs. stereoscopic), navigational device (computer mouse vs. HTC controller), and display type (Tobii pro spectrum 24-inch computer monitor vs. HMD Tobii Pro HTC Vive). For both conditions we used the same high-performance desktop computer with two graphics processing units. Participants navigated through virtual space using a controller or mouse to teleport (a risk free navigation (Keshavarz et al. 2015)) from one point to another. The distance between teleport points was a maximum distance allowed by the HTC Vive controller in order to minimise the navigation effort and allow users to quickly move around the space (see Fig. 2).

For the paper condition, we used the content prepared by architects to be presented to stakeholders involved in the project (investors, various engineers, etc). Since we planned to present this content to future occupants of the building we removed some detailed information intended for engineers. Such information could overwhelm participants and they could miss important information asked in the questionnaires (described below). Together with architects we decided to leave in the plans the measures of the building,

[2] https://innorenew.eu/.

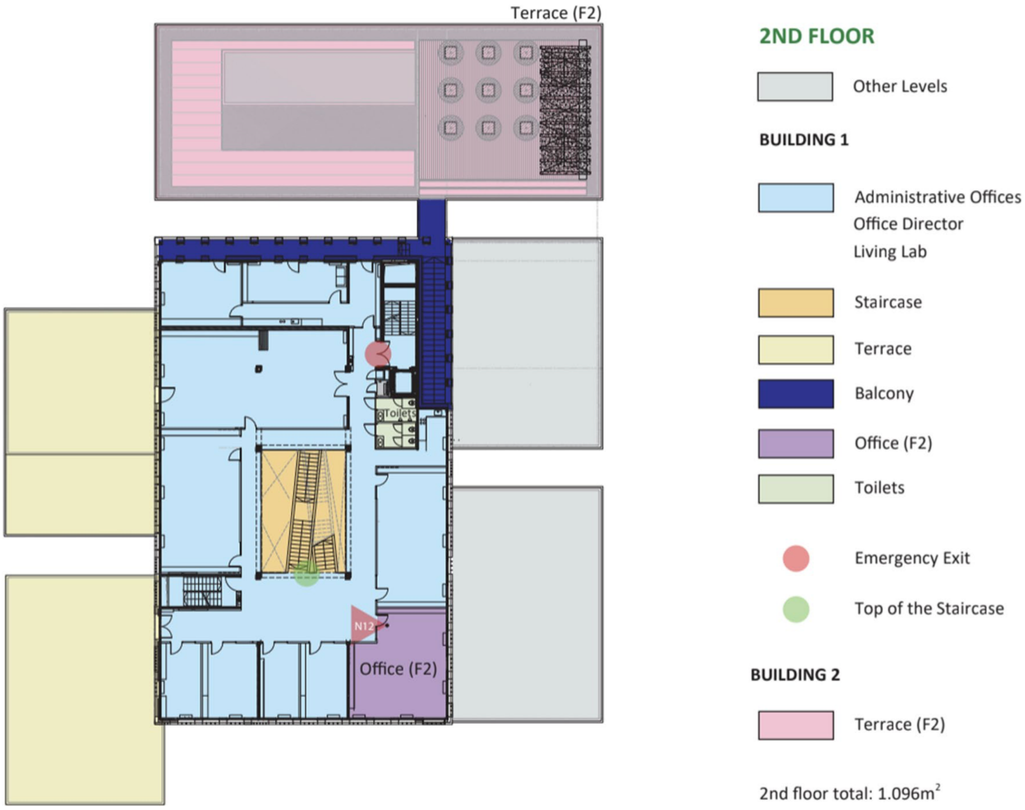


Fig. 3 One of the 2D floor plans from the Innorenew CoE building

names of rooms and areas, and to highlight 12 areas that architects labelled as interesting (see Fig. 3 as an example). We also replaced the original rendered images with the ones from the virtual environment looking at the same scenes from the same angles in order to have the same quality of renders as in the virtual experience.

3.2 Participants

In total, 29 participants volunteered to participate in the study: 10 in paper (4 female), 9 in desktop (5 female), and 10 in VR condition (5 female). All participants were 25 to 55 years old ($\bar{x} = 37$) and were either employed by the InnoRenew CoE (80%) or otherwise linked to the institute. The voluntary response sampling method was used by advertising the study on the internal mailing list of the institute since we wanted to study future occupants of the building.

3.3 Study procedure

Participants were first asked to sign a consent form and fill in a short pre-questionnaire about the current office space and moving to a new building. Participants have been randomly assigned to conditions. After a five minute training session, we calibrated the eye-tracking device and attached the electrodermal activity sensors (EDA or galvanic skin response (GSR)). The data from these sensors are out of scope of this paper and will be published in a separate publication. Participants were then instructed to freely explore the representation method for up to 15 minutes without any restrictions or tasks given in order to imitate how users actually explore a physical building. To facilitate navigation, information boards and signs were included in the environment as in an actual building (the paper condition had labels of rooms and spaces included).

Participants were then given a post-questionnaire about their experience and attitude towards moving, two standard questionnaires – a sense of presence (SUS) (Mel et al. 1998; Usoh and Catena 2000; Usoh et al. 1999) and a user experience questionnaire (UEQ) (Schrepp et al. 2017) – and a questionnaire to assess participants' spatial perception of the environment. These questionnaires were answered without the access to the representation method. Next, participants answered also a size and capacity perception questionnaire

Table 1 Study procedure for each condition

Time	Paper condition	Desktop condition	VR condition
	Consent form & Pre-questionnaire (current building, moving expectations)		
5 min training	Floor plan, rendered images of outdoor of the building with plan reading instructions.	Demo virtual experience with instruction on how to teleport using the mouse.	Demo virtual experience with instruction on how to teleport using the controller.
	Attach EDA sensor on participants' wrist & calibrate the eye-tracker		
Up to 15 min exploration			
	Post-questionnaire (representation experience & moving expectations), SUS, UEQ and Spatial perception questionnaires		
	Size and capacity perception questionnaire with the use of the representation method		
Up to 15 min exploration	VR condition	VR condition	/
	Post-questionnaire (demographics, VR experience, sight problems, sense of orientation)		

with the use and help of the representation method they previously explored. The results of the last two questionnaires will also be published in a separate publication.

After completing the questionnaires, participants assigned to the paper and desktop conditions were invited to explore the virtual building using the HMD VR. At the end, participants were asked to fill in a short post-questionnaire with demographic questions, questions about previous experience with VR technology (3D games or virtual experiences), sight problems and sense of orientation. The overall procedure of the study is shown in Table 1.

3.4 Data collection

In all conditions we captured eye-tracking and EDA data (as mentioned, these will be published in our later work). For paper and desktop conditions facial expression data were also collected to measure affective engagement and consequently user experience (due to the use of HMD we could not do it in VR) (Kapoor et al. 2003). For both desktop and VR conditions, path tracking data were logged by the system as a part of behaviour engagement (for the paper condition this was not possible due to representation specifics). Cognitive engagement was not measured since the study did not involve any task to solve. In all conditions users completed all questionnaires mentioned.

We used Affectiva iMotions (iMotions n.d.) real-time facial expression analysis (FEA) software for the data acquisition and expression analysis with the AFFDEX face detection (Farnsworth 2019). The software is based on the Facial Action Coding System (FACS), which codes specific combinations of action units (contractions of facial muscles) into the six basic emotions (Farnsworth 2019; McDuff et al. 2016): joy, anger, surprise, fear, disgust and sadness. Due to the relevance to the study, we have taken into account only joy and anger, to which we will refer from here on as positive and negative emotions. Affectiva iMotions provides emotion evidence scores corresponding to the probability of the presence of each emotion between 0 (absent) and 100 (present). A threshold suggested in the literature for an emotion being present or absent is between 50 and 70 (Farnsworth 2019). In order to avoid noise, we decided on a threshold for the presence of emotional response at a minimum expression duration of one second (1s), added an immediate median correction of the last three (3) samples of the emotion evidence score, and set the threshold at 70 (Weber 2018; Farnsworth 2019).

To measure the sense of presence we used the common Slater-Usoh-Steed (SUS) six questions questionnaire (Mel et al. 1998; Usoh and Catena 2000; Usoh et al. 1999), which measures: (i) the sense of being in the VR environment, (ii) the extent to which the VR environment becomes the dominant reality, and (iii) the extent to which the VR environment is remembered as a 'place'. For measuring the user experience we used the short version of the User Experience Questionnaire (UEQ-S) (Schrepp et al. 2017; Schrepp 2019) with eight items/questions. The first four represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four hedonic qualities (Stimulation and Novelty) (Schrepp 2019).

3.5 Data analysis

In all statistical analysis we used a significance level $p - value = 0.05$ and a restrictive confidence interval (CI) of 95%. Each data set collected in the study was first checked for normality using the Shapiro-Wilk normality test (Shapiro and Wilk 1965). The statistical significance between study conditions (between-subject design) was examined using the Kruskal-Wallis non parametric test (Kruskal-wallis test 2008), and if significance was found we used the Mann-Whitney test (Neuhäuser 2011) to determine where the significance occurred. The resulting $p < 0.05$ are reported as statistically significant. All boxplots use a 1.5xIQR (interquartile range) rule and Tukey's fences (Lisa 2016) for whiskers and identified outliers. Asterisk notation is used in figures to visualise statistical significance (not significant: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, and ****: $p < 0.0001$).

We also conducted a power analysis to check and validate the results and findings of the study. We calculated the effect size (Cohen's d) for each data set collected (Cohen 1988), selected the minimum effect size (Cohen's $d = 1.553$) and estimated the statistical power ($1 - \beta = 0.9$) of data to check whether the type II error probability (β) is within an acceptable range for a given sample size ($n = 9$ per group) and a significance level ($\alpha = 0.05$). The estimated power value 0.9 shows that with the given sample size, we can have a 90% chance that we correctly reject the null hypothesis with a significance level 0.05.

Since all independent variables (engagement (behavioural, affective), sense of presence, and user experience) are, based on the literature, related to one another (see Sections 2.1 and 2.2) we believe that observing them together can better support the results. We have thus used the Pearson multiple correlation test (Plackett 1983) to find out whether there were any correlations between these variables.

4 Results and findings

This section is divided into four subsections. The first focuses on user engagement, the second on the sense of presence and user experience, the third describes the correlation between these measures, and the fourth subsection presents the results from the pre- and post-questionnaire about moving.

4.1 User engagement results

In this section we present results divided in two parts: behaviour and affective engagement (see Section 2.1).

4.1.1 Behavioural engagement

Behavioural engagement covers observable behaviour, which includes among others conscious navigation, involvement observation, time spent, the amount and type of interaction with the environment, etc. Overall, participants in the paper condition spent on average five (5) minutes to freely explore their representational method compared to 10 in the desktop and 14 in the VR conditions. For further behavioural engagement we turned to observations for the paper condition and logs for the other two conditions. In particular, we looked at the paths traversed and time spent at different interesting areas.

Figure 4 shows the average amount of time spent at each teleport point. In the VR condition, participants navigated through more teleport points and spent more time outside of the building compared to the desktop condition. This happened on *Terrace 1*, *Staircase 1-2* (1st floor), *Terrace 2*, the passage between *Terrace 2* and the building (2nd floor), and in the *Open space* (3rd floor). It thus shows higher engagement in these areas. An interesting observation is that in both desktop and VR conditions participants moved towards the windows wherever this was possible and looked at the view outside. We observed this in the *meeting room* on the first floor, in the *office room* on the second floor, and in the *open space* on the third floor, which is visible by circles close to the windows.

Figure 5 illustrates the navigation patterns and density of the average time in seconds for the overall experience for desktop and VR conditions. In the VR condition, participants explored the outdoors (the side of the building with the entrance including the road) a lot more compared to desktop condition. There is no big difference in the indoor navigation pattern between the conditions. Nevertheless, participants in VR showed higher behavioural engagement (more time spent in the environment, more teleportation points explored, more time spent in more interesting areas) compared to the desktop condition. In the paper condition, participants mostly just scanned the plans and spent more time on rendered images. Comparing the overall time, however, they spent twice as less time in the desktop and three times in the VR condition.

4.1.2 Affective engagement measures

We did a facial expression analysis for the paper and desktop conditions to determine the emotional engagement of participants as described in Sect 3.4. Since the duration of exploration in each condition and the number of emotions

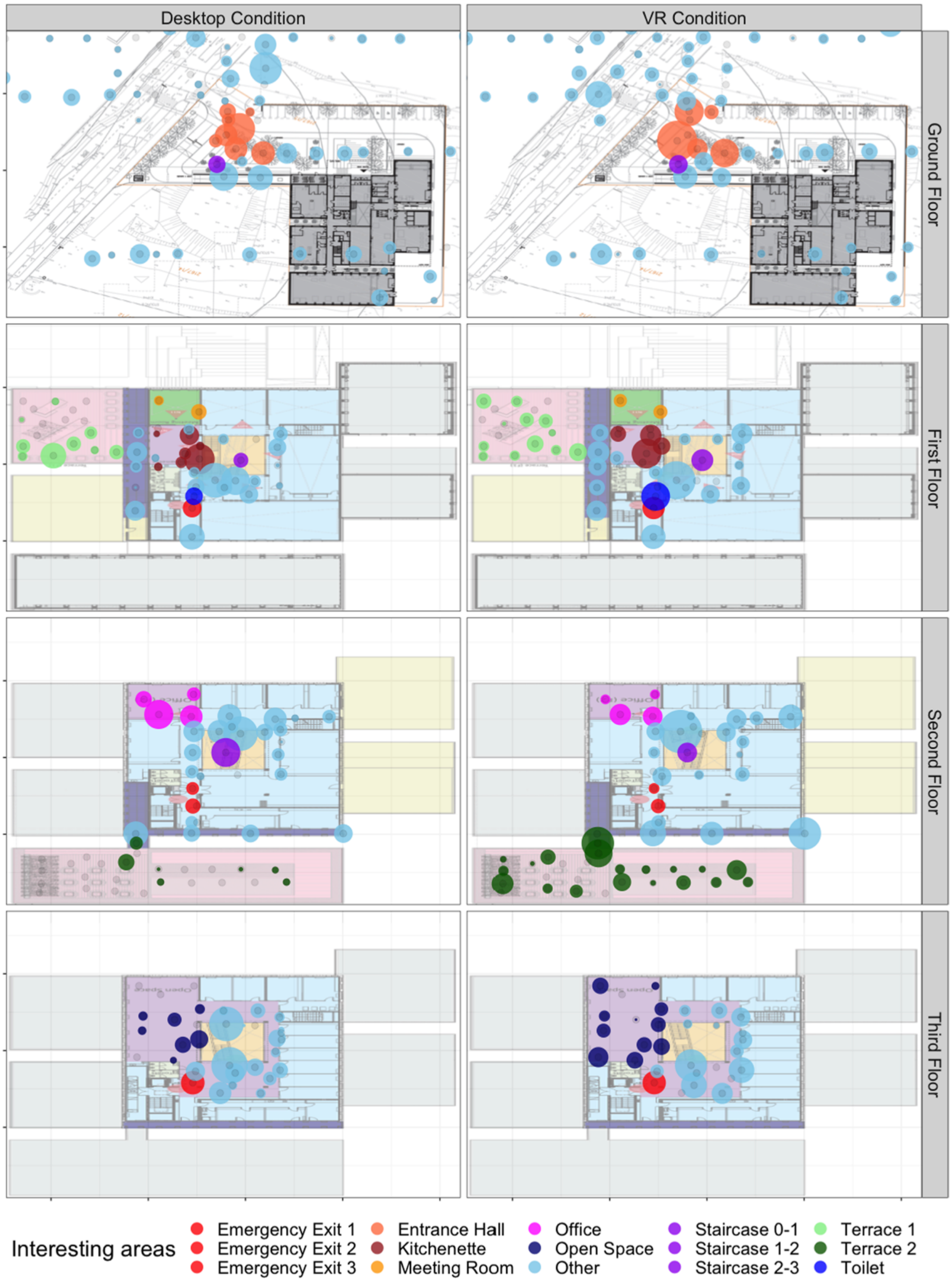


Fig. 4 The average amount of time in seconds spent at each teleport point in each interesting area for desktop and VR conditions. The bigger the circle, more time spent at that point

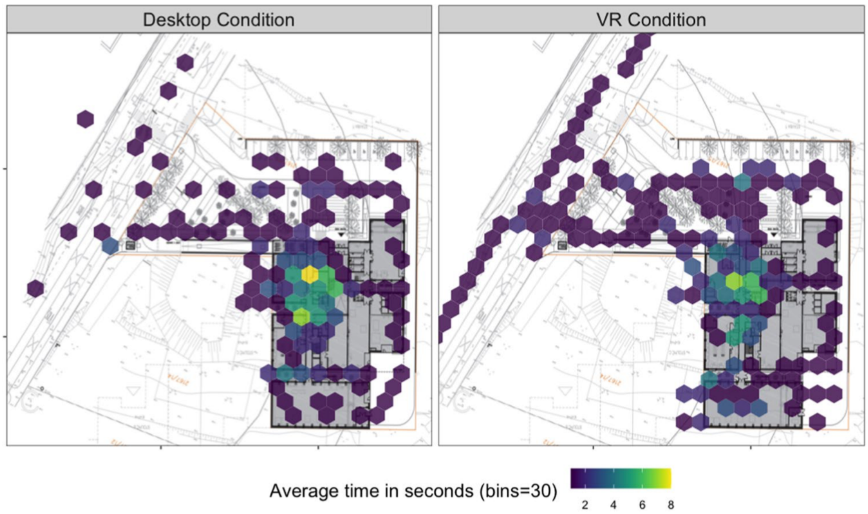


Fig. 5 Navigation pattern and density of the average time in seconds for the overall experience presented on a hexbin map

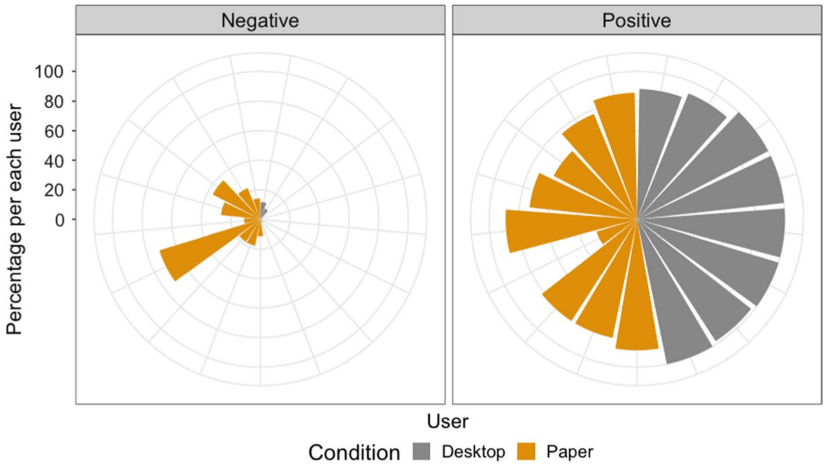


Fig. 6 Percentage of positive and negative emotions per user in paper and desktop conditions. The longer the cone, the more emotional responses of that kind were registered

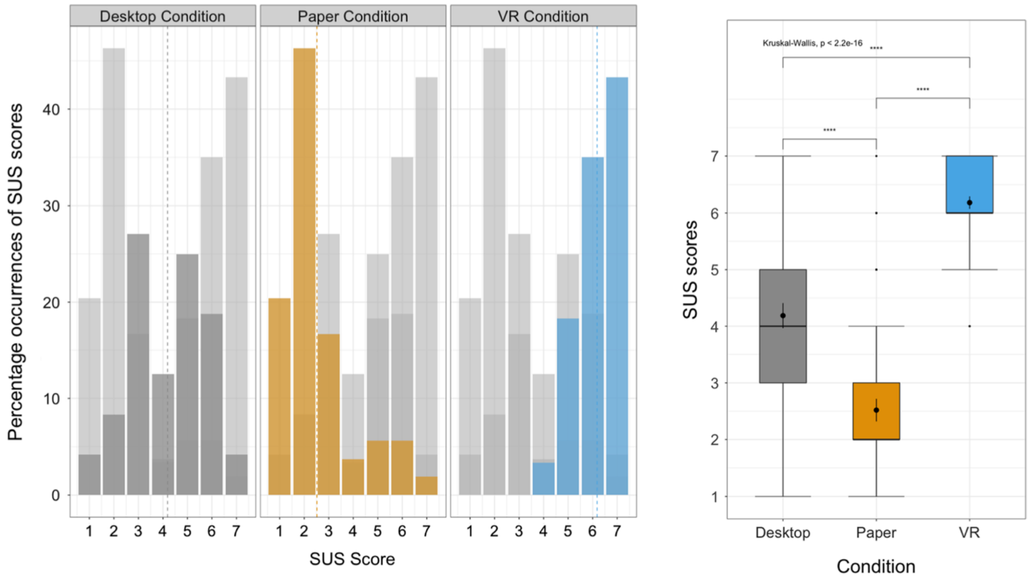
registered varied between participants, we present the values in percentages – this is the number of each emotion registered out of all the emotions registered for a specific participant. Fig. 6 shows a slight presence of negative emotions in the paper condition and more positive emotions registered in the desktop condition.

While the mean percentage of registered positive emotion instances is significantly higher than negative emotion instances in both conditions, it is even higher in the desktop compared to the paper condition (desktop $\bar{x} = 97.3\%$, CI [94.6, 100]; paper $\bar{x} = 74.4\%$, CI [64, 84.8]). In the paper condition, the system registered nearly 25% of negative emotion instances, while this number is low in the desktop condition. The Mann-Whitney nonparametric test shows a significant difference in emotions registered for both positive and negative emotions when comparing desktop and paper conditions (positive $p = .0065$, negative $p = .0065$).

Table 2 Descriptive statistics for both emotions considered together (positive and negative). The percentage difference from the overall average in the last column was calculated as such $100 * (Total\_average - Study\_condition\_average)/Total\_average$

Study Condition	Average for both emotions	SD	Error 95 CI	% difference from overall average
Desktop	25.0	13.7	7.9	+ 10.5%
Paper	30.5	19.6	10.8	− 9.4%
Total	27.9	16.9	6.8	

Descriptive statistics for both emotions are summarised in Table 2. High standard deviations indicate the widespread of emotions. Consequently, the mean value is a poor indicator of individual performance, but considering the entire data it gives and indication on the variations in



(a) Distribution of mean values of SUS answers on a 7-point scale for each condition.

(b) Study condition influence on emotions registered.

Fig. 7 SUS scores

emotions registered between study conditions. The Mann-Whitney nonparametric test did not show statistical significance ($W = 111.5$, $p = 0.2682$)

4.2 Sense of presence and user experience

This sub-section reports on the results from the Slater-Usoh-Steed (SUS) and the User Experience Questionnaire (UEQ).

4.2.1 Sense of presence (SUS) questionnaire

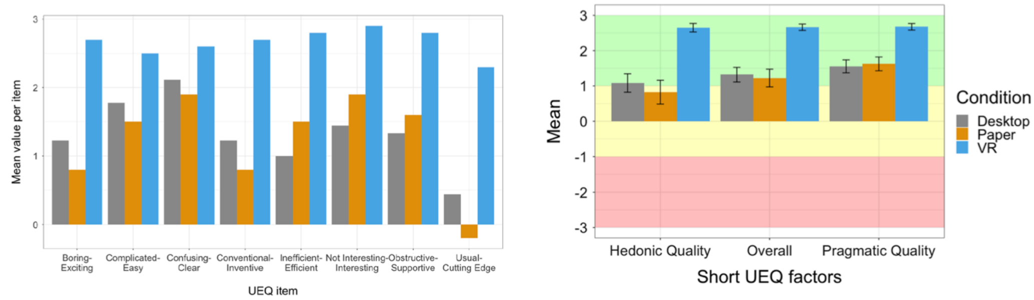
The answers to the six SUS questions are reported on a 7-stage Likert scale (1 low, 7 high). Figure 7a illustrates the mean values from the SUS questionnaire and distribution. The answers for the VR condition are skewed towards the left with the mean value > 6, which means that in this condition the sense of presence was very high compared to the other two conditions. The distribution of the paper condition is skewed towards the right with the mean value < 3, which means that the sense of presence was lowest here. The sense of presence for the desktop condition had a more varied distribution with a mean value of around 4.

A Kruskal-Wallis nonparametric test showed a significant difference ($\chi^2 = 93.531$, df = 2, $p = 2.2e - 16$), indicating that the condition influenced the sense of presence. The pairwise comparison using a Mann-Whitney-Wilcoxon nonparametric test (desktop-paper: $p = 1.7e - 07$, desktop-VR: $p = 4.5e - 11$ and paper-VR: $p = 1.4e - 16$) revealed significant differences between all the pairwise study conditions (Fig. 7b).

4.2.2 User experience questionnaire

The answers to UEQ questions/items are also reported on a 7-stage Likert scale, where the first four items represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four the hedonic qualities (Stimulation and Novelty). The results are assigned a value between −3 and 3 as follows: if the item starts with a negative term, a 1 on the 7-stage scale becomes a −3, while if the item starts with a positive term, the 1 becomes a +3. These adjustments are kept throughout the rest of the analysis.

The mean values are calculated and rated according to (Schrepp et al. 2017; Schrepp 2019). Every UEQ scale belongs to either a 'negative evaluation' with a mean value $-3 \leq \bar{x} < -0.8$, a 'neutral evaluation' with a mean value between -0.8 and +0.8, and a 'positive evaluation' with a mean value $+0.8 < \bar{x} \leq +3$. This can be applied to single items or factors (pragmatic and hedonic). The calculated mean for every item is shown in Fig. 8a. There is only one negative evaluation for the paper condition (for the question usual-cutting edge) and no negative evaluation for desktop and VR conditions.



(a) Mean values per item/question.

(b) Mean values of UEQ factors (pragmatic and hedonic) and all item/question together (overall).

Fig. 8 UEQ scores

Figure 8b shows the mean values per factor and the corresponding confidence intervals. In addition, the mean value of all the items is given as an overall user experience value. Pragmatic ($\bar{x} = 2.67$) and hedonic ($\bar{x} = 2.65$) qualities, as well as overall user experience ($\bar{x} = 2.66$) are perceived as extremely positive in the VR condition. The mean values of all the factors are also clearly above the threshold value of 0.8 in the desktop condition (pragmatic - $\bar{x} = 1.55$, hedonic - $\bar{x} = 1.08$ and overall - $\bar{x} = 1.31$). The mean of the hedonic factor ($\bar{x} = 1.31$), which represents the Stimulation and Novelty, is below the threshold value of 0.8 in the paper condition.

4.3 Correlations between user experience, sense of presence and engagement

A Pearson multiple correlation test (Plackett 1983) was used to find out if there were any correlations between user experience, sense of presence, and engagement (behavioural, affective). We found correlations between *user experience – sense of presence* ($r = 0.72$, $p = 0.009$) and *user experience – behavioural engagement* ($r = 0.69$, $p = 0.010$) in the VR condition. There were no significant correlations between user experience, sense of presence, and engagement for desktop and paper conditions.

4.4 Attitude towards moving

In the pre-questionnaire, participants were asked about the current office premises and about the attitude towards moving into the new building. Only 4% of participants rated the current premises as bad, 11% as excellent, 83% as good and 2% do not know. The reasons mentioned against the current premises were: not enough light in the office (in the old city centre buildings are in close proximity), air quality, sharing an office with other people, improvised offices and lack of parking spaces. There were more reasons to like the offices such as convenient location, close proximity to the sea side, wooden furniture, plenty of light, creative teammates, and good equipment.

The pre-questionnaire also revealed some concerning comments about the new building being located in a different town. Most of the participants like the current town as it is bigger and it offers a variety of facilities in close proximity, which is not the case with the new building. More than half of the participants complained that they will need to commute to the new location as they live near the current one. Just one person mentioned that they live close to the new building and one liked its remote location. One person raised the fact that their children are going to school in the current town and the move will require a bit of flexibility from the family. A few mentioned that they will miss seeing researchers from the university and two assumed they will be split between the new and the old location, which would, in their opinion, make meetings a challenge.

When asked in a pre-questionnaire about how they expect the new office building will compare to the current one 93% of the responses claimed it will be better (3.5% the same and 3.5% do not know). Positive reasons included dedicated offices and labs, and that the whole company will be together as they are residing in several buildings at the moment. These results did not significantly change in the post-questionnaire.

When asked in the post-questionnaire whether the expectations of moving to a new building changed 90% of participants in the VR condition answered yes, compared to 37.5% in desktop and 66.7% in the paper condition (see Table 3). The positive attitude in VR could also be seen in rating the VR experience as excellent in 90% of the cases (50% in desktop and 44% in paper).

Table 3 Post-questionnaire results of quantitative questions

How would you rate the experience of the future InnoRenew office facilities?					
Condition	**Poor**	**Below average**	**Above average**	**Excellent**	**I don't know**
VR			10%	90%	
Desktop			50%	50%	
Paper			44%	44%	12%
Did this experience change your expectations about how your future office facilities will look like?					
Condition	**Yes**	**No**	**I don't know**		
VR	90%	10%			
Desktop	38%	62%			
Paper	67%	33%			
How confident would you feel about moving around in a normal working day?					
Condition	**Lost**	**Confident**	**Very confident**	**I don't know**	
VR		30%	70%		
Desktop		50%	50%		
Paper	12%	44%	44%		

Participants in the VR condition provided 22.5% more answers to open ended questions than in desktop and 32.5% more than in the paper condition. In, VR participants provided 16 positive comments about offices and the building, 6 complemented greenery and surroundings (including parking), and two close proximity to colleagues (11, 1, and 2 in desktop and 8, 1, and 2 in paper condition, respectively). The expanded list of positive reasons for moving included: modernity, new equipment, wooden furniture, spacious common areas, natural sunlight, flexibility, eco-friendliness, clean design, cosy and pleasant environment, working together with colleagues, and availability of parking spaces. Regarding the confidence of moving about in the new facilities, positive comments focused on how well participants understood the space, and in particular participants provided 10 positive comments in VR, 7 in desktop and 5 in the paper condition. Only in the latter we received comments such as *"I felt lost"* and *"through the maps it's hard to understand the locations"*, which points to the fact the allocentric view is not an optimal representation for some people.

It could be argued that regardless of the representation method, participants who were offered more information about the future office building are likely to find more positive reasons for moving. However, participants in desktop and VR conditions were exposed to the same detailed virtual environment and yet, in the VR condition more comments in general were received. Another reason might be the time spent in the VR condition (on average 40% more) and thus exposing participants to more information. But the results show that in both conditions participants visited the same places and used similar traversing patterns. Overall, the VR condition has elevated a more positive attitude towards moving.

5 Discussion

We designed an explorative study to compare how future occupants are able to project themselves into the future office space when presented with a different representation method of the real future building: (i) a paper-based presentation with floor plans and rendered images of the building as used by architects (paper condition), (ii) a highly-detailed 3D virtual model of the future building experienced on a computer monitor (desktop condition), and (iii) the same virtual environment experienced with a head mounted display (VR condition). The impact of the representation method was measured through user engagement, sense of presence and user experience.

5.1 User engagement

Our RQ1a was *How does the representation method of the model of the future office space influence engagement?* Based on (Appleton et al. 2008), engagement can be divided into affective, cognitive and behavioural. We hypothesised that the representation method will affect engagement.

5.1.1 Behavioural engagement

We measured behavioural engagement for desktop and VR conditions by observing where users spent their time and how much time they spent in different locations. We could not observe this in the paper condition; nevertheless, we could observe users being more interested in renders (egocentric view) than in floor plans (allocentric view). This was expected as floor plans can be hard to mentally translate to an egocentric view for non professionals (Kuliga and James 2020), which has also been found in this study. In desktop and VR conditions participants spent a similar amount of time in different spaces with some exceptions. In the VR condition, participants spent more time in front of the building exploring the road to and from the building, and on outdoor terraces. In both conditions, participants also spent a considerable amount of time at the windows admiring the view.

It is well known that individual and architectural factors can directly and indirectly contribute to physical and psychological discomfort in an office space (Alan 1989). The time spent staring out of the windows and observing the views should not come as a surprise, since attractive window views are known for reducing discomfort at work (van Esch 2019) and consequently improving home life (e.g. sleep) (Aries et al. 2010), provide a natural source of light (Wong 2017), while closer proximity to windows reduces health problems and complaints among occupants (Küller and Wetterberg 1996; Yildirim et al. 2007).

A similar explanation can be found for spending time in outdoor spaces (terraces and in front of the entrance). These spaces are filled with greenery, which provides an opportunity for recovery from mental fatigue and are generally beneficial to human health (Berman et al. 2008; Kaplan 1995). The time spent in the VR condition "walking" to and from the building can be attributed to higher immersion with the environment surrounding users. Moving in space has been pointed out as one of the pivotal affordances in VR (Zhao et al. 2020; Horvat et al. 2019) and an important finding of our study was that people spent nearly 40% more time moving in VR compared to the desktop. There is a potential to explore these observations in further studies.

5.1.2 Affective engagement

We measured affective engagement through analysis of facial expression for desktop and paper conditions. The amount of positive feelings was higher in the desktop condition. Significantly more negative feelings were recorded in the paper condition, while these were almost absent in the desktop condition. Affective engagement is directly linked to the sense of presence and emotions. For example, experience in anxious or relaxing VR environments increased these emotions as well as the sense of presence (Riva 2007). Studies on learning in VR also report on higher emotional arousal in VR compared to content presented on a computer monitor (Makransky et al. 2019; Parong and Mayer 2020). Higher affective engagement of the desktop condition can be explained by the pleasant environment presented by the aforementioned greenery (Berman et al. 2008; Kaplan 1995). We could not measure affective engagement in the VR condition; although, since the virtual environment was the same in the desktop condition as in the VR condition, and since users spent considerably more time in the VR condition, we can assume that the VR condition would show even higher affective engagement based on the increased behavioural engagement and sense of presence.

5.2 Sense of presence and user experience

Research questions RQ1b and RQ1c were *How does the representation method influence the sense of presence and user experience?* We hypothesised, based on previous studies, that the more immersive the environment the higher the level of the sense of presence. We measured the sense of presence with the SUS questionnaire. As expected, users in the VR condition reported the highest sense of presence, while users in the paper condition the lowest. The significant difference has been confirmed between all three study conditions. This was not surprising for the paper condition since it does not offer an immersive experience (Juan et al. 2018). Despite experiencing the same virtual environment, the significant difference between desktop and VR conditions should not come as a surprise since several studies already confirmed that compared to desktop, VR supports a higher sense of presence (Makransky et al. 2019; Zhao et al. 2020; Federica et al. 2019).

As both desktop and VR conditions used the same highly-detailed virtual environment, we expected a lesser difference between the two. This is because prior work in architecture claimed that the more detailed the environment, the higher the sense of presence (Bertol 1996); however, these works did not mention the representation method. Another reason was that our study just involved casual exploration of a future building compared to other studies examining different representation methods (VR vs desktop) that involved complex tasks such as playing a video game (Federica et al. 2019) or navigating and understanding a non-detailed virtual environment (Ruddle et al. 1999).

Despite expectations, our study shows that even if users do not engage in complex tasks, and they just move around a virtual building, there was a significant difference between the desktop and the VR representation method. This is in line with aforementioned studies (Federica et al. 2019; Ruddle et al. 1999) and it also indicates that the sense of presence might not be so much related to the task at hand but more on the level of detail (Bertol 1996) and immersivness provided by the display.

In our study, only the display (high definition computer monitor vs. head-mounted display (HMD)) facilitated this difference. Previous studies have shown that HMDs can enhance focused and selective attention (Cho et al. 2002; Amprasi et al. 2022), and consequently increase the sense of presence, by cutting visual stimuli from the physical environment. To even further leverage the sense of presence researchers have explored small additions such as a breeze or adjusted temperature in the physical surroundings (Ranasinghe et al. 2017). It would be interesting to investigate if such additions would increase the sense of presence even for the desktop condition. In addition, since previous studies report a lower sense of presence with teleportation vs walking, future studies could also look at this aspect.

User experience in VR has been often looked at in relation to the sense of presence or even defined by it (Steuer 1992). It was thus expected that the paper condition would receive least positive values. Despite using the same virtual environment for desktop and VR, the VR condition acquired twice the levels of positive scoring for both hedonic and pragmatic qualities of user experience. This finding further supports the results for a higher sense of presence in VR.

5.3 Correlation between measures

We have also investigated how different variables are correlated and how they could be combined in order to more accurately present engagement, sense of presence, and user experience for each study condition. This contrasts to the existing body of VR research in architectural design that often looks at these aspects individually. Our results show correlations in VR between *the sense of presence and user experience*, which was to be expected since the two are considered related (Steuer 1992), as well as between *user experience and behavioural engagement*. Interestingly, there were no correlations found in the other two conditions. Despite using the same environment in desktop and VR, the latter supports significantly higher degree of immersion. It thus highly affects the sense of presence, which affected the user experience. However, further studies are required to explore any causation between the aspects measured.

5.4 Implications for research and practice

A recent review of the VR landscape in architecture and construction defined six areas where VR supports the field: stakeholder engagement, design support, design review, construction support, operations and management support, and training (Delgado et al. 2020). Our study focuses on under explored aspects of stakeholder engagement. VR has already been used to engage with various stakeholders (investors, potential clients, public), giving them an opportunity to participate in pre-occupancy evaluation, examine built-assets at real-scale or align their expectations with the actual design (Pejic et al. 2017; Juan et al. 2018; Kini and Shilpa 2019; Frost and Peter 2000).

However, the focus of our study was on a particular group of stakeholders – the future occupants. As the company leadership already decided to move to new premises, this group did not have much say in this, contrasting our work with other studies of future occupants, which focused on VR as a support of better understanding of, and the decision-making process concerning their future workplace (Westerdahl 2006; Frost and Peter 2000). For this reason we were interested in how involving future occupants in the process by showing them different representation methods (paper, desktop, VR) would help them project themselves into their future office space and how this might influence their attitudes towards moving. In our study we have shown that VR as a representation method increases the sense of presence, user experience and engagement and that these measures are correlated.

The study also shows that for this particular group of users, the building itself is not the only important part of the experience. The exterior of the building as well as the greenery around it proved to be almost as essential. Participants enjoyed moving outdoors and observing the building from various sides. Several even took "a walk" to the building, which is one of the key aspects of VR (Jiayan and Sensibaugh 2020; Horvat et al. 2019). Moreover, the majority of participants stopped at the windows and admired the view. Overall, the VR experience influenced participants to add more positive comments about moving, compared to other conditions. Even participants that have initially experienced a less immersive representation method (paper and desktop) provided positive comments after experiencing the VR. These comments further emphasise the importance of feedback for people in leadership positions.

The following results are speculative due to the explorative nature of the study presented but point out interesting research directions. It has been shown that past experiences support imagining hypothetical future scenarios and that memory can flexibly recombine past and novel experiences into novel simulations of possible future events (Szpunar 2010; Schacter 2012; Pascal 2008; Suddendorf and Corballis 2007). The ability to imagine details of future scenarios also supports one's effective coping with future events through emotional regulation and appropriating current activities (Brown 2002). VR might thus provide a viable approach in supporting "one's effective coping with future events" such as a job relocation – a life event often associated with stress (Martin 2000)). Previous findings also show that imagining future scenarios is associated with optimism bias (Tali 2011). Besides other measures used, the amount of additional positive comments received after experiencing the VR condition might thus also indicate that VR better supports simulating possible futures, compared to the other two representation methods. These speculative results open up possible new research directions in different fields to further investigate the effects of VR on projecting oneself into future scenarios. For example, it has been shown that representing potential future selves in a virtual environment can motivate people to make better food choices in the future (Kuo et al. 2016).

The main VR challenges listed in recent literature include insufficiently detailed virtual environments, the necessity of supervision and hardware, the ease of use, and the sense of isolation (Delgado et al. 2020; Juan et al. 2018). In contrast with older studies with future occupants (Westerdahl 2006; Frost and Peter 2000), we have built a highly-detailed virtual environment, used high-end hardware and teleportation as navigation techniques to avoid any technical and other issues that might contribute to user discomfort. None of the participants reported any problems related to VR. However, such a project needs a large time commitment for implementation (Yung and Khoo-Lattimore 2019) (4 months of work in our case), which might limit the capability of implementing such an environment for a lot of projects as it involves significant financial investment. The claims that VR environments can cut the costs of time spent on decision-making and the cost of building physical mock-ups (Majumdar et al. 2006; Juan et al. 2018) remain largely unproven as VR is not yet part of the essential toolchain in the fields of architecture, engineering, and construction. In spite of all the benefits presented in this study and other studies mentioned in the paper, VR remains useful and nice to have but not essential, yet.

We also did not encounter any issues with the ease of use, despite the fact that 87% of our participants have never played 3D video games, and 64% never experienced VR before. As for the supervision, hardware, and sense of isolation, they still remain a challenge. Although some attempts in this direction have been made (Du 2018; Kunert 2020) the current systems still do not allow for large scale collaborations in VR. Nevertheless, as technology advances such limitations might be resolved with a broader commercial uptake.

5.5 Limitations and future research directions

One of the main limitations of our study is the number of participants. At the time of the study the Innorenew CoE institute had 39 people employed. Some were not eligible to participate in the main study (architects and people in leadership positions). The majority of other employees have participated and observations were consistent among each condition. Nevertheless, we have shown that the type II error is within the acceptable range. This paper does also not discuss how different methods of representation affects the spatial perception as this is beyond the scope of the topic presented here. Such an analysis, results and discussion is the subject of future work.

6 Conclusion

The focus of the explorative study presented in this paper was to investigate how different representation methods (VR, desktop, paper) influenced how future occupants project themselves into the future office space and to explore their willingness to relocate. The results show a statistically significant difference for the sense of presence, user experience and user engagement in VR compared to the paper and desktop conditions. Users were also observed looking at the views through the windows, spent time on terraces surrounded by the greenery, extensively explored the surroundings, and even "took a walk" to work. We also received more positive comments about the building after experiencing the VR condition. We argue that experiencing VR better supported people in projecting themselves into their future office spaces, increased their capability to imagine future scenarios, and positively affected their attitude towards moving. The study, experimental design, results and discussion presented here can inform future studies and aid the current development of systems for exploring future buildings.

Funding The authors acknowledge the European Commission for funding the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and European Union of the European Regional Development Fund). We also acknowledge support from the Slovenian research agency ARRS (program no. P1-0383, J1-9186, J1-1715, J5- 1796, J1-1692, and IO-0035).

Code availability Not applicable

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethics approval The authors declare that they received the ethics approval to conduct the research.

Consent to participate The authors declare that they received the consent from all the participants before participating in the study.

References

Acock Malcolm (1985) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. In: By David Marr. The Modern Schoolman 62(2):141–142. https://doi.org/10.5840/schoolman198562219

Amprasi Evaggelia et al. (2022) Effect of a Full Immersive Virtual Reality Intervention on Selective Attention in Children. In: International Journal of Instruction 15.1

Maike Andresen, Margenfeld Jil (2015) International relocation mobility readiness and its antecedents. In: J Managerial Psycho 30(3):234–249. https://doi.org/10.1108/JMP-11-2012-0362

Appleton James J, Christenson Sandra L, Furlong Michael J (2008) Student engagement with school: Critical conceptual and methodological issues of the construct. In: Psychol the Schools 45(5):369–386. https://doi.org/10.1002/pits.20303

Aries Myriam BC, Veitch Jennifer A, Newsham Guy R (2010) Windows, view, and officce characteristics predict physical and psychological discomfort. J environ psychol 30(4):533–541. https://doi.org/10.1016/j.jenvp.2009.12.004

Azhar Salman (2011) Building information modeling (BIM): Trends, benefits risks, and challenges for the AEC industry. Leadersh manag eng 11(3):241–252

Bakker Niels H, Passenier Peter O, Werkhoven Peter J (2003) Effects of Head-Slaved Navigation and the Use of Teleports on Spatial Orientation in Virtual Environments. Human Factors 45(1):160–169. https://doi.org/10.1518/hfes.45.1.160.27234

Banos Rosa M et al (2004) Immersion and Emotion: Their Impact on the Sense of Presence. CyberPsychol Beh 7(6):734–741. https://doi.org/10.1089/cpb.2004.7.734

Barfield Woodrow, Weghorst Suzanne (1993) The sense of presence within virtual environments: A conceptual framework. Adv Human Factors Ergon 19:699–699

Bassanino May et al. (2010) The impact of immersive virtual reality on visualisation for a design review in construction. In: Proceedings of 2010 14th International Conference Information Visualisation. IEEE, 585–589

David Benyon et al (2014) Presence and digital tourism. AI & society 29(4):521–529. https://doi.org/10.1007/s00146-013-0493-8

Berg Leif P, Vance Judy M (2017) An industry case study: Investigating early design decision making in virtual reality. In: Journal of Computing and Information Science in Engineering 17(1):011001-1-011001-7. https://doi.org/10.1115/1.4034267

Berman Marc G, John Jonides, Stephen Kaplan (2008) The cognitive benefits of interacting with nature. In: Psychological science 19.12, 1207–1212. https://www.jstor.org/stable/40064866

Daniela Bertol (1996) Designing digital space: an architect's guide to virtual reality. John Wiley & Sons, New York City

Birenboim Amit et al (2019) The utilization of immersive virtual environments for the investigation of environmental preferences. Landsc Urban Plan 189:129–138. https://doi.org/10.1016/j.landurbplan.2019.04.011

Bowman Doug A, David Koller, Larry F Hodges (1997) Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques.In: Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality. IEEE, 45–52. https://doi.org/10.1109/VRAIS.1997.583043

Pascal Boyer (2008) Evolutionary economics of mental time travel? Trends in Cognit Sci 12(6):219–224. https://doi.org/10.1016/j.tics.2008.03.003

Brade Jennifer et al (2017) Being there againpresence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. Int J Hum Comput Stud 101:76–87. https://doi.org/10.1016/j.ijhcs.2017.01.004

Brown Gary P et al (2002) Worry and the simulation of future outcomes. Anxiety, Stress Coping 15(1):1–17. https://doi.org/10.1080/10615800290007254

Busch Marc et al. (2014) Being there for real: presence in real and virtual environments and its relation to usability. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, 117–126. https://doi.org/10.1145/2639189.2639224

Cartillier Vincent et al. (May 2021) Semantic MapNet: Building Allocentric Semantic Maps and Representations from Egocentric Views. In: Proceedings of the AAAI Conference on Artificial Intelligence 35.2, 964–972.https://ojs.aaai.org/index.php/AAAI/article/view/16180

Castronovo Fadi et al. (2013). An evaluation of immersive virtual reality systems for design reviews. In: Proceedings of the 13th international conference on construction applications of virtual reality, 22–29.https://itc.scix.net/pdfs/convr-2013-2.pdf

Cebeci Berk, Celikcan Ufuk, Capin Tolga K (2019) A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments. Comp Anim Virtual Worlds 30(3–4):e1893. https://doi.org/10.1002/cav.1893

Chandrasegaran Senthil K et al (2013) The evolution, challenges, and future of knowledge representation in product design systems. Computeraided des 45(2):204–228. https://doi.org/10.1016/j.cad.2012.08.006

Chevi Oriane (2018) The role and challenges for the HR department in corporate relocation context. Nova School of Business and Economics(NSBE). http://hdl.handle.net/10362/39724

Cho Baek Hwan et al. (2002) Attention enhancement system using virtual reality and EEG biofeedback. In: Proceedings IEEE Virtual Reality 2002.IEEE, 156–163

Cohen Jacob (1988) Statistical power analysis for the behavioral sciences.Academic press

Dajana Snopková et al (2019) Navigation in Indoor Environments: Does the Type of Visual Learning Stimulus Matter?In: ISPRS Int. J GeoInf 8(6):251. https://doi.org/10.3390/ijgi8060251

Delgado Juan MD et al. (2020) A research agenda for augmented and virtual reality in architecture, engineering and construction.In: Advanced Engineering Informatics 45, p. 101122.https://doi.org/10.1016/j.aei.2020.101122

Dey Sanorita et al. (2014) Study the Effect of Allocentric and Egocentric Interfaces in Indoor Navigation. In: of the HCI Consortium. HCI Consortium, 1–6

Du Jing et al (2018) Zero latency: Real-time synchronization of BIM data in virtual reality for collaborative decision-making. Autom Constr 85:51–64. https://doi.org/10.1016/j.autcon.2017.10.009

Dunston Phillip S et al. (2011) An immersive virtual reality mock-up for design review of hospital patient rooms. In: Collaborative design in virtual environments. Springer, 167–176. https://doi.org/10.1007/978-94-007-0605-7_15

Eby Lillian T, Russell Joyce E (2000) Predictors of employee willingness to relocate for the firm. J Vocat Behav 57(1):42–61. https://doi.org/10.1006/jvbe.1999.1724

Semiha Ergan et al (2019) Quantifying human experience in architectural spaces with integrated virtual reality and body sensor networks. J Comput Civil Eng 33(2):04018062. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000812

van Esch Emmy et al (2019) Office window views: View features trump nature in predicting employee well-being. J Environ Psychol 64:56–64. https://doi.org/10.1016/j.jenvp.2019.05.006
Faas Daniela et al (2014) The in uence of immersion and presence in early stage engineering designing and building. AI EDAM 28(2):139–151. https://doi.org/10.1017/S0890060414000055
Farnsworth Bryn (2019) Facial Action Coding System (FACS) - A Visual Guidebook.https://imotions.com/blog/facial-action-coding-system/(visited on 11/12/2020)
Fernando Terrence P, Kuo-ChengWu, May Bassanino (2013) Designing a novel virtual collaborative environment to support collaboration in design review meetings. In: Journal of Information Technology in Construction 18, 372–396. http://www.itcon.org/paper/2013/19
Frantzidis Christos A et al (2010) Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. IEEE transact Inf Technol Biomed 14(3):589–597. https://doi.org/10.1109/TITB.2010.2041553
Fredricks Jennifer et al. (2011) Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. In: Regional Educational Laboratory Southeast. https://files.eric.ed.gov/fulltext/ED514996.pdf
Frost Peter, Peter Warren (2000) Virtual reality used in a collaborative architectural design process. In: Proceedings of IEEE 2000 Conference on Information Visualization. An International Conference on Computer Visualization and Graphics, 568–573. https://doi.org/10.1109/IV.2000.859814
Gifford Robert (2007) Environ psychol Princ pract. Optimal books Colville, WA
Marc Hassenzahl (2010) Experience design: Technology for all the right reasons. Synth lect human-centered inf 3(1):1–95. https://doi.org/10.2200/S00261ED1V01Y201003HCI008
Alan Hedge et al (1989) Work-related illness in offices: a proposed model of the sick building syndrome. Environ Int 15(1–6):143–158. https://doi.org/10.1016/0160-4120(89)90020-2
Heydarian Arsalan et al. (2014). Immersive virtual environments: experiments on impacting design and human building interaction. In: Proceedings of CAADRIA 2014 19th International Conference on the Association of Computer-Aided Architectural Design Research in Asia, 729–738.http://papers.cumincad.org/data/works/att/caadria2014_161.content.pdf
Christoph Hölscher et al (2006) Up the down staircase: Wayfinding strategies in multi-level buildings. J Environ Psychol 26(4):284–299. https://doi.org/10.1016/j.jenvp.2006.09.002
Horvat Nikola et al. (2019) Comparing virtual reality and desktop interface for reviewing 3D CAD models. In: Proceedings of the Design Society: International Conference on Engineering Design. Vol. 1. 1. Cambridge University Press, 1923–1932. https://doi.org/10.1017/dsi.2019.198
iMotions (n.d.). Facial Expression Analysis. Version 2018.https://imotions.com/biosensor/fea-facial-expression-analysis/
ISO (2010). ISO 9241-210:2010 Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. (visited on 11/29/2021)
ISO (2018) ISO 19650-1:2018 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling – Part 1: Concepts and principles.https://www.iso.org/standard/68078.html(visited on 12/08/2021)
Jaeger Beverly K, Ronald R Mourant (2001) Comparison of Simulator Sickness Using Static and Dynamic Walking Simulators. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 45. 27. SAGE Publications Sage CA: Los Angeles, CA, 1896–1900.https://doi.org/10.1177/154193120104502709
Juan Yi-Kai, Chen Hsing-Hung, Chi Hao-Yun (2018) Developing and evaluating a virtual reality-based navigation system for pre-sale housing sales. Appl Sci 8(6):952. https://doi.org/10.3390/app8060952
Kaplan Stephen (1995) The restorative benefits of nature: Toward an integrative framework. J environ psychol 15(3):169–182. https://doi.org/10.1016/0272-4944(95)90001-2
Kapoor Ashish, Yuan Qi, Rosalind W Picard (2003) Fully automatic upper facial action recognition. In: Proceedings of IEEE 2003 International SOI Conference (Cat. No. 03CH37443). IEEE, 195–202.https://doi.org/10.1109/AMFG.2003.1240843
Katsis Christos D et al (2008) Toward emotion recognition in car-racing drivers: A biosignal processing approach. IEEE Trans Sys , Man, and Cybern Part A: Sys Humans 38(3):502–512. https://doi.org/10.1109/TSMCA.2008.918624
Kearsley Greg, Shneiderman Ben (1998) Engagement theory: A framework for technology-based teaching and learning. Educ technol 38(5):20–23
Keshavarz Behrang et al. (2015) Vection and visually induced motion sickness: how are they related? In: Frontiers in Psychology 6, p. 472. issn:1664-1078. https://doi.org/10.3389/fpsyg.2015.00472.https://www.frontiersin.org/article/10.3389/fpsyg.2015.00472
Kini Vishwas G, Shilpa Sunil (2019) Design and realization of sustainable rural housing using immersive virtual reality platform. In: Proceedings of 25th ACM Symposium on Virtual Reality Software and Technology, 1–2. https://doi.org/10.1145/3359996.3364719
Klein Stanley B (2013) The complex act of projecting oneself into the future. Wiley Interdisciplinary Rev Cognit Sci 4(1):63–79. https://doi.org/10.1002/wcs.1210
Koelstra Reinder AL (2012) Affective and Implicit Tagging using Facial Expressions and Electroencephalography. Queen Mary University of London
Krukar Jakub, Ruth C Dalton, Christoph Hölscher (2016) Applying HCI Methods and Concepts to Architectural Design (Or Why Architects Could Use HCI Even If They Don't Know It). In: Architecture and Inter- action: Human Computer Interaction in Space and Place. Ed. by Nicholas S. Dalton et al. Cham: Springer International Publishing, 17–35. isbn:978-3-319-30028-3.https://doi.org/10.1007/978-3-319-30028-3_2
Kruskal-wallis test (2008) In: The concise encyclopedia of statistics. New York, NY: Springer New York, 288–290. isbn: 978-0-387-32833-1.https://doi.org/10.1007/978-0-387-32833-1_16
Kuliga Saskia, James Charlton, et al. (2020) Developing a Replication of a Wayfinding Study. From a Large-Scale Real Building to a Virtual Reality Simulation. In: German Conference on Spatial Cognition. Springer, 126–142. https://doi.org/10.3929/ethz-b-000418825
Kuliga Saskia, Thrash Tyler et al (2015) Virtual reality as an empirical research tool-Exploring user experience in a real building and a corresponding virtual model. Comput Environ Urban Syst 54:363–375. https://doi.org/10.1016/j.compenvurbsys.2015.09.006
Küller Rikard, Wetterberg Lennart (1996) The subterranean work environment: impact on well-being and health. Environ Int 22(1):33–52. https://doi.org/10.1016/0160-4120(95)00101-8
Kunert Andre et al (2020) Multi-Window 3D Interaction for Collaborative Virtual Reality. IEEE Trans Visual Comput Gr 26(11):3271–3284. https://doi.org/10.1109/TVCG.2019.2914677
Kuo Hsu-Chan, Lee Chun-Chia, Chiou Wen-Bin (2016) The power of the virtual ideal self in weight control: Weight-reduced avatars can enhance the tendency to delay gratification and regulate

dietary practices. Cyberpsycho Beh Soc Netw 19(2):80–85. https://doi.org/10.1089/cyber.2015.0203
Law Effie LC et al. (2007) Towards a UX manifesto. In: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference, 1–2.https://doi.org/10.14236/ewic/HCI2007.95
Leon Marianthi et al (2014) Development and Testing of a Design Protocol for Computer Mediated Multidisciplinary Collaboration during the Concept Stages with Application to the Built Environment. Proc Environ Sci 22:108–119. https://doi.org/10.1016/j.proenv.2014.11.011
Lisa Sullivan (2016) InterQuartile Range (IQR).http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_summarizingdata/bs704_summarizingdata7.html#headingtaglink_3(visited on 01/08/2020)
Llorach Gerard, Alun Evans, Josep Blat (2014) Simulator Sickness and Presence Using HMDs: Comparing Use of a Game Controller and a Position Estimation System. In: Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology. Edinburgh, Scotland: Association for Computing Machinery, 137–140. https://doi.org/10.1145/2671015.2671120
Luo Lu, Cooper Cary L (1990) Stress of job relocation: Progress and prospect. Work & Stress 4(2):121–128. https://doi.org/10.1080/02678379008256974
Majumdar Tulika, Martin A Fischer, Benedict R Schwegler (2006) Conceptual design review with a virtual reality mock-up model. In: Joint international conference on computing and decision making in civil and building engineering, 2902–2911.http://www.irbnet.de/daten/iconda/CIB21113.pdf
Makransky Guido, Terkildsen Thomas S, Mayer Richard E (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn Instr 60:225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007
Robin Martin (1999) Adjusting to job relocation: Relocation preparation can reduce relocation stress. J Occup Organ Psychol 72(2):231–235. https://doi.org/10.1348/096317999166626
Martin Robin et al (2000) The role of attributions in psychological reactions to job relocation. Work & Stress 14(4):347–361. https://doi.org/10.1080/02678370010029186
McDuff Daniel et al. (2016) AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 3723–3726. https://doi.org/10.1145/2851581.2890247
Mel Slater, John McCarthy, Francesco Maringelli (1998) The Inuence of Body Movement on Subjective Presence in Virtual Environments. Human Factors 40(3):469–477. https://doi.org/10.1518/001872098779591368
Meng Zhang et al. (2020) Chapter 1 - Digital twin driven smart product design framework. In: Digital Twin Driven Smart Design. Academic Press, 3–32. https://doi.org/10.1016/B978-0-12-818918-4.00001-4
Moghimi Mohammadhossein et al (2016) Inuencing human affective responses to dynamic virtual environments. Presence Teleoperators and Virtual Environ 25(2):81–107. https://doi.org/10.1162/PRES_a_00249
Murugappan Muthusamy et al. (2008) Time-frequency analysis of EEG signals for human emotion detection. In: Proceedings of 4th Kuala Lumpur international conference on biomedical engineering 2008. Springer, 262–265.https://doi.org/10.1007/978-3-540-69139-6_68
NBS (2016) What is Building Information Modelling (BIM)?https://www.thenbs.com/knowledge/what-is-building-information-modelling-bim(visited on 12/28/2019)
Neuhäuser Markus (2011) Wilcoxon-Mann-Whitney Test. Ed. by Miodrag Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 1656–1658.isbn: 978-3-642-04898-2. https://doi.org/10.1007/978-3-642-04898-2_615
Federica Pallavicini, Pepe Alessandro, Minissi Maria E (2019) Gaming in Virtual Reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games? Simul Gaming 50(2):136–159. https://doi.org/10.1177/1046878119831420
Parong Jocelyn, Mayer Richard E (2020) Cognitive and affective processes for learning science in immersive virtual reality. J Comput Ass Learning 37(1):226–241. https://doi.org/10.1111/jcal.12482
Pashler Harold E (1999) The psychology of attention. MIT press
Pejic Petar et al. (2017) Application of Augmented and Virtual Reality in Residential Complex Presentation, Case Study: Energoprojekt Sunnyville.In: Journal of Industrial Design & Engineering Graphics 12.1.https://doi.org/10.5937/fmet1702227P
Pierce Matthew B, Philip A Young, Shawn M Doherty (2017) Engagement and Competence in VR and non-VR Environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2082–2085.https://doi.org/10.1177/1541931213602003
Plackett Robin L (1983) Karl Pearson and the chi-squared test. In:International Statistical Review/Revue Internationale de Statistique, 59–72.https://doi.org/10.2307/1402731
Ranasinghe Nimesha et al. (2017) Ambiotherm: enhancing sense of presence in virtual reality by simulating real-world environmental conditions. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 1731–1742. https://doi.org/10.1145/3025453.3025723
Regenbrecht Holger T, Schubert Thomas W, Friedmann Frank (1998) Measuring the Sense of Presence and its Relations to Fear of Heights in Virtual Environments. Int J Human-Comput Interact 10(3):233–249. https://doi.org/10.1207/s15327590ijhc1003_2
Renner Rebekka S, Velichkovsky Boris M, Helmert Jens R (2013) The perception of egocentric distances in virtual environments-a review. ACM Comput Surveys (CSUR) 46(2):1–40. https://doi.org/10.1145/2543581.2543590
Riva Giuseppe et al (2007) Affective interactions using virtual reality: the link between presence and emotions. CyberPsychol Behav 10(1):45–56. https://doi.org/10.1089/cpb.2006.9993
Ruddle Roy A, Payne Stephen J, Jones Dylan M (1999) Navigating Large-Scale Virtual Environments: What Differences Occur Between Helmet-Mounted and Desk-Top Displays? Presence 8(2):157–168. https://doi.org/10.1162/105474699566143
Sagie Abraham, Krausz Moshe, Weinstain Yehuda (2001) To move or not to move: Factors affecting employees' actual relocation when an entire plant moves. J Occup Organ Psychol 74(3):343–358. https://doi.org/10.1348/096317901167398
Sanchez-Vives Maria V, Slater Mel (2005) From presence to consciousness through virtual reality. Nat Rev Neurosci 6:332–339. https://doi.org/10.1038/nrn1651
Schacter Daniel L et al (2012) The future of memory: remembering, imagining, and the brain. Neuron 76(4):677–694. https://doi.org/10.1016/j.neuron.2012.11.001
Schaufeli Wilmar B (2013). What is engagement. In: Employee engagement in theory and practice. Vol. 15. Routledge London, 29–49
Schnabel Marc A, Kvan Thomas (2003) Spatial understanding in immersive virtual environments. Int J Archit Comput 1(4):435–448. https://doi.org/10.1260/147807703773633455
Schrepp Martin (2019) UEQ-User Experience Questionnaire.https://www.ueq-online.org/ (visited on 01/08/2020)
Schrepp Martin, Hinderks Andreas, Thomaschewski Jörg (2017) Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). Int J Interact Multimed Artif Intell 4(6):103–108. https://doi.org/10.9781/ijimai.2017.09.001

Schwind Valentin et al. (2019) Using Presence Questionnaires in Virtual Reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, 1–12.https://doi.org/10.1145/3290605.3300590

Shapiro Samuel S, Wilk Martin B (1965) An Analysis of Variance Test for Normality (Complete Samples). In: Biometrika 52(3/4):591–611http://www.jstor.org/stable/2333709

Tali Sharot (2011) The optimism bias. Curr biol 21(23):941–945. https://doi.org/10.1016/j.cub.2011.10.030

Slater Mel, Martin Usoh (1993) Presence in immersive virtual environments. In: Proceedings of IEEE Virtual Reality Annual International Symposium. IEEE, 90–96. https://doi.org/10.1109/VRAIS.1993.380793

Slater Mel, Usoh Martin, Steed Anthony (1995) Taking Steps: The In uence of a Walking Technique on Presence in Virtual Reality. ACM Transact Comput Human Interact 2(3):201–219. https://doi.org/10.1145/210079.210084

Steuer Jonathan (1992) Defining virtual reality: Dimensions determining telepresence. J commun 42(4):73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Suddendorf Thomas, Corballis Michael C (2007) The evolution of foresight: What is mental time travel, and is it unique to humans? Behav Brain Sci 30(3):299–313. https://doi.org/10.1017/S0140525X07001975

Szpunar Karl K (2010) Episodic future thought: An emerging concept. Perspect Psychol Sci 5(2):142–162. https://doi.org/10.1177/1745691610362350

Takahashi Kazuhiko, Akinori Tsukaguchi (2003) Remarks on emotion recognition from multi-modal bio-potential signals. In: Proceedings of SMC'03 2003 IEEE International Conference on Systems, Man and Cy-bernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483). Vol. 2. IEEE, 1654–1659. https://doi.org/10.1109/ICIT.2004.1490720

Topu Fatma B, Goktas Yuksel (2019) The effects of guided-unguided learning in 3d virtual environment on students' engagement and achievement. Comput Human Behav 92:1–10. https://doi.org/10.1016/j.chb.2018.10.022

Usoh Martin, Kevin Arthur, et al. (1999) Walking > Walking-in-Place > Flying, in Virtual Environments. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH'99. USA: ACM Press/Addison-Wesley Publishing Co., 359–364.https://doi.org/10.1145/311535.311589

Usoh Martin, Catena Ernest et al (2000) Using Presence Questionnaires in Reality. Presence Teleoperators and Virtual Environ 9(5):497–503. https://doi.org/10.1162/105474600566989

Violante Maria G, Vezzetti Enrico, Piazzolla Pietro (2019) How to design a virtual reality experience that impacts the consumer engagement:the case of the virtual supermarket. Int J Interact Des Manuf (IJIDeM) 13(1):243–262. https://doi.org/10.1007/s12008-018-00528-5

Weber Marlene (2018) Automotive emotions: a human-centred approach towards the measurement and understanding of drivers' emotions and their triggers. Brunel University London

Weber Marlene et al (2019) Investigation of the dependency of the drivers' emotional experience on different road types and driving conditions. Transp res part F traffic psychol behav 65:107–120. https://doi.org/10.1016/j.trf.2019.06.001

Westerdahl Börje et al (2006) Users' evaluation of a virtual reality architectural model compared with the experience of the completed building. Autom Constr 15(2):150–165. https://doi.org/10.1016/j.autcon.2005.02.010

Witmer Bob G, Singer Michael J (1998) Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence Teleoperators and Virtual Environ 7(3):225–240. https://doi.org/10.1162/105474698565686

Wong Liang (2017) A review of daylighting design and implementation in buildings. Renew Sustain Energy Rev 74:959–968. https://doi.org/10.1016/j.rser.2017.03.061

Xie Haiyan, Eduard Tudoreanu, Wei Shi (2006) Development of a virtual reality safety-training system for construction workers. In: Digital library of construction informatics and information technology in civil engineering and construction, 1–9

Yazdani Ashkan, Jong-Seok Lee, Touradj Ebrahimi (2009) Implicit emotional tagging of multimedia using EEG signals and brain computer interface.In: Proceedings of the first SIGMM workshop on Social media, 81–88. https://doi.org/10.1145/1631144.1631160

Yildirim Kemal, Akalin-Baskaya Aysu, Celebi Mine (2007) The effects of window proximity, partition height, and gender on perceptions of openplan offices. J Environ Psychol 27(2):154–165. https://doi.org/10.1016/j.jenvp.2007.01.004

Yung Ryan, Khoo-Lattimore Catheryn (2019) New realities: a systematic literature review on virtual reality and augmented reality in tourism research. Current Issues Tour 22(17):2056–2081. https://doi.org/10.1080/13683500.2017.1417359

Zanbaka Catherine A et al (2005) Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. IEEE Trans Visualization and Comput Gr 11(6):694–705. https://doi.org/10.1109/TVCG.2005.92

Zhao Jiayan, Peter LaFemina, et al. (2020) Learning in the Field: Comparison of Desktop, Immersive Virtual Reality, and Actual Field Trips for Place-Based STEM Education. In: Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 893–902.https://doi.org/10.1109/VR46266.2020.00012

Jiayan Zhao, Sensibaugh Tesalee et al (2020) Desktop versus immersive virtual environments: effects on spatial learning. Spatial Cognition Comput 20(4):328–363. https://doi.org/10.1080/13875868.2020.1817925

A Spatial Understanding

The spatial understanding of the environment was analysed in terms of participants' abilities of estimating distance and directions.

A.1 Distance Estimations

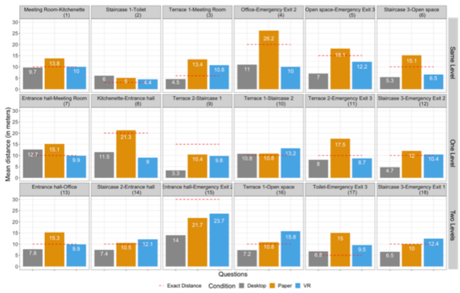


Fig. 9: Means of estimated distances in metres between 14 interesting areas for each question and exact distances marked with the red dashed line for each question.

Figure 9 shows an overestimation of most distance estimations (approximately 70% of the questions) for the paper condition and underestimation of most distance estimations (approximately 70% of the questions) for desktop condition.

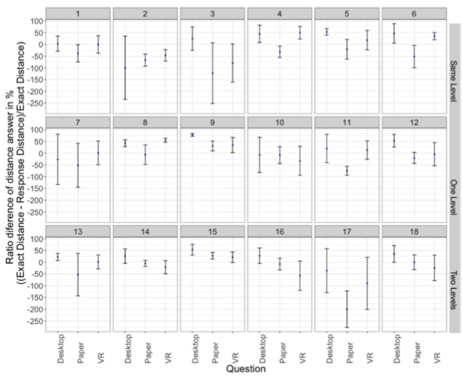


Fig. 10: Mean error ratios of distance estimations for each question. A positive value indicates that participants underestimated the distances (exact distance > estimated distance) while a negative value means that participants overestimated the distance (exact distance < estimated distance). Error bars show 95% CIs.

Figure 10 shows mean error ratios for each question. Participants' distance estimations in VR condition were closer to exact (real) values (lower error ratio) for the most questions compared to other two conditions.

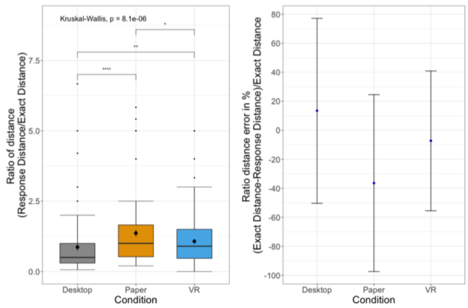


Fig. 11: Left: study condition influence on distance estimations. Right: ratio between correct distance and the estimated answers (a positive value indicates that participants underestimated the distances (exact distance > estimated distance) while a negative value means that participants overestimated the distance (exact distance < estimated distance). Error bars show 95% CIs.

Previous findings show that users underestimate egocentric distance values in non-immersive as well as immersive virtual environments by approximately 15% Thompson et al. 2004. The results in our study show a significant difference between all pairs of conditions. While users in paper conditions had all building and room sizes available in their presentation (floor plans, rendered images), users in desktop and VR condition did not. Nevertheless, users in paper condition overestimated distances as well as contributed to the highest error rate (36%). Users in desktop and VR conditions underestimated distances and users in VR condition were best in distance estimation with the error rate of around 7% (desktop 13%). While the results are better that in previous studies, the users in VR were twice as best. The highly-detailed virtual environment offering several depth cues probably lead to the results comparable to real world distance estimations Bob G Witmer and Sadowski Jr 1998; Thompson et al. 2004. Active exploration of the virtual environment can also lead to accurate egocentric distance between virtual objects Thompson et al. 2004.

A.2 Direction Estimations

To analyse the direction estimation, all the answers were converted to eight cardinal directions. Each cardinal direction was assigned a value between 0 and 7 (clockwise) relative to the participant's position and direction they were facing (facing direction is the north). These adjustments are kept throughout the rest of the analysis.

As Figure 12 shows, there was no considerable pattern present in any of the three study conditions when it comes to estimating direction. All the study groups show deviations from the exact (correct) direction estimation for most of the direction questions

Figure 13 right reports mean error ratios for each condition. On average, participants' direction estimation in VR

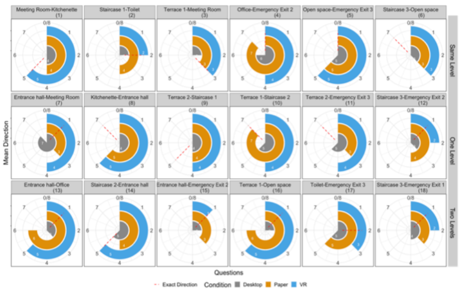


Fig. 12: Means of estimated directions between defined interesting areas with respect to the floor level difference between interesting areas (*same level*, *one level* difference and *two levels difference*). The exact directions are marked with the red dashed lines for each question. The direction scale from 0-7 is defined in text.

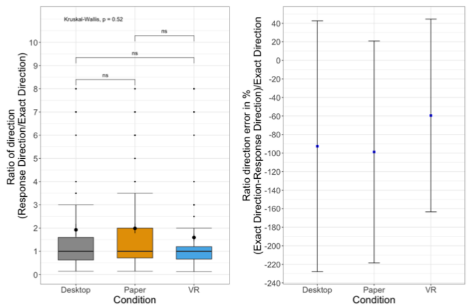


Fig. 13: Left: the influence of the study condition on estimating the direction to an interesting area from participants' position and direction they are facing. Right: ratio between correct direction and the estimated answers. Error bars show 95% CIs.

condition showed a lower error ratio (direction error = -63.0%, CI [-166.9,40.9]) compared to other two conditions. While participants' direction estimation in paper condition showed a higher error ratio (direction error = -102.3%, CI [-221.5, 16.9]), which is nearly twice as in VR condition. The high standard error indicates the widespread of direction estimation responses found between participants.

The findings in the literature show approximately 85% correctness in direction estimation while observing simple objects in non-realistic desktop based virtual environment Albert, Rensink, and Beusmans 1999. Other studies showed that estimating directions in real, high-detailed desktop based and VR based environment are comparable and amount to around 8° error for adults Waller, Beall, and Loomis 2004; Jansen-Osmann and Fuchs 2006. While estimations using direction circles had an error twice as high Waller, Beall, and Loomis 2004. It has to be stresses out that in all these studies participants had clear tasks to perform before being testes. The free-style tour used in our experiment clearly affected our results.

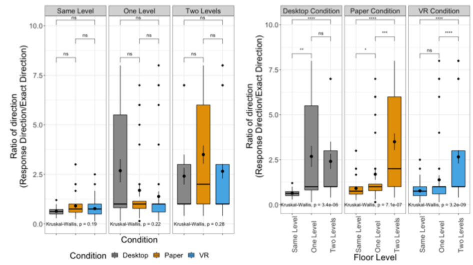


Fig. 14: Left: study condition influence on direction estimations and floor level difference between interesting areas (*same level*, *one level* difference and *two levels* difference). Right: floor level difference between interesting areas influence on direction estimation for each condition.

B Volumetric Understanding

The volumetric understanding of the environment was analysed in terms of participants' abilities of estimating dimensions of the functional space.

B.1 Size Estimations

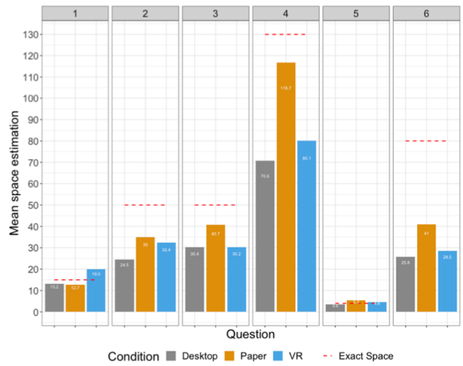


Fig. 15: The mean values of participants' size estimations for interesting areas compared to exact area marked with a red dashed line.

Figure 15 shows an underestimation of most area size estimations (approximately 85% of the questions) for all three conditions.

Figure 16 reports mean error ratios for each question. Participants' size estimations in paper condition were closer to exact (real) values (lower error ratio) for the most of the questions compared to other two conditions. While participants' size estimations in desktop condition were more likely further from exact (real) values (higher error ratio) compared to VR condition.

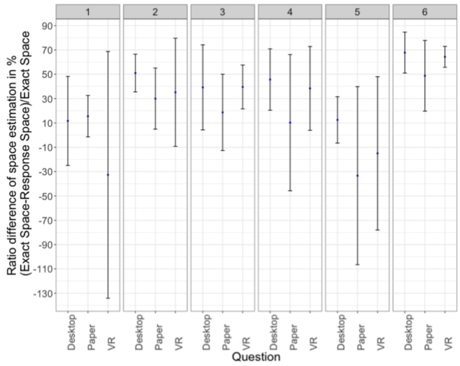


Fig. 16: Mean error ratios of participants' size estimations for each question. A positive value indicates that participants underestimated the the area size (exact size > estimated size) and the negative value means that participants overestimated the area size (exact size < estimated size). Error bars show 95% CIs

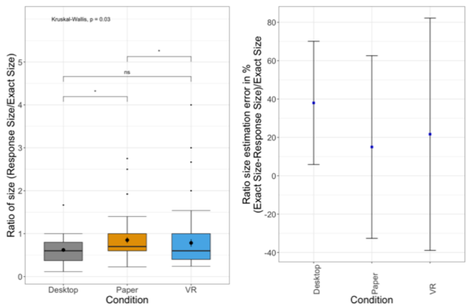


Fig. 17: Left: study condition influence on size estimations. Right: ratio between correct size and the estimated answers (a positive value indicates that participants underestimated the area size (exact size > estimated size) while negative value means that participants overestimated the area size (exact size < estimated size). Error bars show 95% CIs.

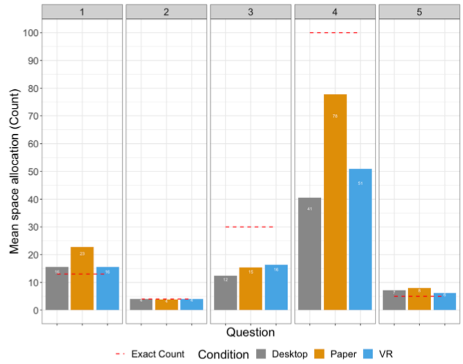


Fig. 18: Participants' mean counts of people that would comfortably fit in a defined interesting areas against exact count marked with a red dashed line, per each question

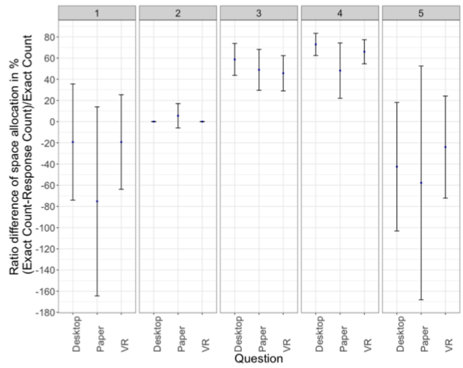


Fig. 19: Mean error ratios of participants' capacity estimations for each question. A positive value indicates that participants underestimated the capacity of a specific interesting area (exact count>response count) and negative value means that participants overestimated the capacity of a specific interesting area (exact count<response count). Error bars show 95% CIs.

Figure 17 right reports mean error ratios for each condition. On average, participants' size estimation in paper condition showed a lower error ratio (error = 14.9%, CI [-35.2,65.0]) compared to other two conditions. While participants' size estimation in desktop condition showed a higher error ratio (error = 37.9%, CI [6.0, 69.8]). This shows that participants in the paper condition in overall had better estimated the sizes compared to other two conditions.

B.2 Capacity Estimations

Figure 18 reports an underestimation for larger spaces and slight overestimation for smaller spaces of most size estimations in all three conditions.

Figure 19 shows mean error ratios for each question. Participants' capacity estimations in paper condition were more likely further from exact (real) values (higher error ratio) for the most of the questions compared to other two conditions. While participants' capacity estimations in VR condition were closer to exact (real) values (lower error ratio) compared to desktop condition.

Figure 20 right reports mean error ratios for each condition. On average, participants' capacity estimation in paper condition showed a lower error ratio (error = -6.1%, CI [-85.8,73.6.0]) compared to other two conditions.

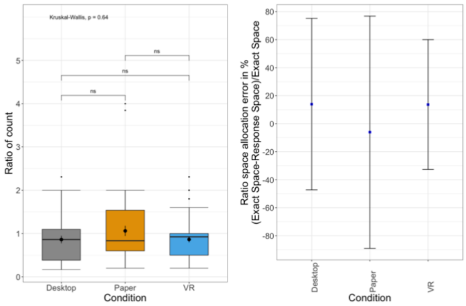


Fig. 20: Left: study condition influence on capacity estimations. Right: ratio between correct count and the estimated answers (a positive value indicates that participants underestimated the capacity of an interesting area (exact count > estimated count) while negative value means that participants overestimated the capacity of an interesting area (exact count < estimated count). Error bars show 95% CIs.

Chapter 5

Publication 02: Arigatō [1]

Title: Arigatō: Effects of Adaptive Guidance on Engagement and Performance in Augmented Reality Learning Environments

Authors: Maheshya Weerasinghe, Aaron Quigley, Klen Čopič Pucihar, Alice Toniolo, Angela Miguel, Matjaž Kljun

Year: 2022

Journal: IEEE Transactions on Visualization and Computer Graphics (Q1)

DOI: 10.1109/TVCG.2022.3203088

Link: `https://ieeexplore.ieee.org/document/9873961`

Summary: The Arigatō Study, aimed to investigate the influence of the *Experiential Learning* cycle and the different adaptation factors of *Guidance* on the user engagement, and learning outcomes in a language learning task. We developed a HMD based AR language learning system that offered immediate assistance to guide learners through different stages of the ExL cycle. This study directly addressed RQ1 (How can XR technology be used to create immersive and interactive experiential learning environments by exploring the importance and effectiveness of different configurations of XR, ExL and instructional guidance?) and then, by designing the amount and type of guidance provided, we addressed RQ3 (How can XR technology be used to create immersive and adaptive experiential learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators?). Arigatō also helped identify XR interventions that can be used to enhance instructional guidance (considered in RQ2). Furthermore, the study investigated how the dynamic adaptations of guidance influenced user experience, engagement, and performance in a language learning task within an immersive ExL environment, which contributed to addressing RQ4 (How do the different configurations of XR and/or degree of dynamic adaptations of guidance affect the user experience, engagement and, consequently, the learning outcomes of immersive ExL environments?).

Arigatō: Effects of Adaptive Guidance on Engagement and Performance in Augmented Reality Learning Environments

Maheshya Weerasinghe, *Student Member, IEEE*, Aaron Quigley, Klen Čopič Pucihar, Alice Toniolo, Angela Miguel and Matjaž Kljun

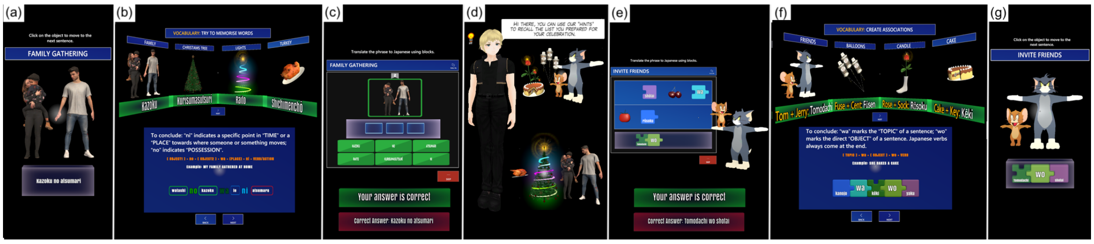


Fig. 1. Arigatō prototype and its experiential learning stages for two different topics: Christmas and birthday celebration. (a) Reflective observation (RO), (b) Abstract conceptualisation (AC), and (c) Active experimentation (AE), stages for fixed-associations condition on the Christmas celebration topic. (d) Concrete experience (CE) for both topics. (e) Active experimentation (AE), (f) Abstract conceptualisation (AC), and (g) Reflective observation (RO) stages for adaptive-associations condition on a birthday celebration topic. To avoid excessive clutter of the AR scene the black background was selected.

Abstract— Experiential learning (ExL) is the process of learning through experience or more specifically "learning through reflection on doing". In this paper, we propose a simulation of these experiences, in Augmented Reality (AR), addressing the problem of language learning. Such systems provide an excellent setting to support "adaptive guidance", in a digital form, within a real environment. Adaptive guidance allows the instructions and learning content to be customised for the individual learner, thus creating a unique learning experience. We developed an adaptive guidance AR system for language learning, we call Arigatō (Augmented Reality Instructional Guidance & Tailored Omniverse), which offers immediate assistance, resources specific to the learner's needs, manipulation of these resources, and relevant feedback. Considering guidance, we employ this prototype to investigate the effect of the amount of guidance (fixed vs. adaptive-amount) and the type of guidance (fixed vs. adaptive-associations) on the engagement and consequently the learning outcomes of language learning in an AR environment. The results for the amount of guidance show that compared to the adaptive-amount, the fixed-amount of guidance group scored better in the immediate and delayed (after 7 days) recall tests. However, this group also invested a significantly higher mental effort to complete the task. The results for the type of guidance show that the adaptive-associations group outperforms the fixed-associations group in the immediate, delayed (after 7 days) recall tests, and learning efficiency. The adaptive-associations group also showed significantly lower mental effort and spent less time to complete the task.

Index Terms—Experiential learning, instructional guidance, adaptive learning systems, augmented reality, engagement, language Learning

1 Introduction

Experiential learning (ExL) is a well-known learning approach used in education, training, facilitation, coaching and organisational development [28, 52, 57]. ExL refers to the process of learning through experience or "learning by doing". One of the most influential models in ExL [57] is the Kolb's learning model [50]. This model defines ExL as "the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience" [50]. The model is represented by four cyclical learning stages as seen in Fig. 2: (1) concrete experience (feeling), (2) reflective observation (watching), (3) abstract conceptualisation (thinking), and (4) active experimentation (doing). One can enter the learning cycle at any stage, but all stages in the cycle must be addressed for meaningful learning to occur.

In learning, including experiential learning, guidance is beneficial. The term guidance has a broad meaning, but we adopt the generally understood definition of, guidance as "any form of assistance offered to users so they can achieve a learning goal". In more precise terms, the role of guidance is "to simplify, provide a view on, elicit, supplant, or prescribe the scientific reasoning skills" [55] involved in pursuing a goal. It can be explicit, as when a teacher provides instructions on how to solve a mathematical problem with all the in-depth explanations of concepts and skills students need to learn, or provides feedback on the current performance. Or it can be implicit, as it is provided within the environment for users to find, interpret and use in order to progress in the task at hand [7, 16]. In either case the amount and the type of guidance play an essential role in engagement and consequently, in the learning outcomes [17, 44, 96].

While several educational definitions of guidance exist [5, 19, 55], in the context of this study we focus on instructional "guidance", which is defined as "providing users with accurate and complete procedural information (and related declarative knowledge) that they have not

- *Maheshya Weerasinghe is with the University of Primorska, Slovenia and the University of St Andrews, United Kingdom. E-mail: amw31@st-andrews.ac.uk.*
- *Aaron Quigley is with the University of New South Wales, Australia. E-mail: a.quigley@unsw.edu.au.*
- *Alice Toniolo and Angela Miguel are with the University of St Andrews, United Kingdom. E-mail: {a.toniolo | arm14}@st-andrews.ac.uk.*
- *Klen Čopič Pucihar and Matjaž Kljun are with the University of Primorska, Slovenia. E-mail: {klen.copic | matjaz.kljun}@famnit.upr.si.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

yet learned in a demonstration about how to perform the necessary sequence of actions and make the necessary decisions to accomplish a learning task" [19]. Guidance also has different dimensions such as "how much" (amount) instructional support it provides [64, 95, 105], and "what kind" (type) of support it provides [5]. Digital technologies such as augmented reality (AR) appear ideal to support adaptation of guidance. AR can provide an environment for simulating various experiences by combining the physical environments with in-context adaptive digital elements that can be manipulated by the user. AR can thus replicate the real-world ExL scenario (e.g. learning by decorating a room for Christmas celebrations with AR objects as can be done with real objects) and show information in a coherent and meaningful way within the real world context.

Researchers have previously explored AR systems as tools to guide or teach different skills and activities that can be learned best through experience [37, 67]. Most of these projects have primarily focused on creating simulated environments to support different stages of the Kolb's experiential learning cycle [12, 37, 67, 68, 97]. However, these projects explored generic guidance only, in which all the users receive the same set of general instructional cues, with the same amount, in the same way [12,41,60,63,101]. Therefore, there are several questions that remain unanswered, such as: How to design and implement guidance in learning environments where the learning system needs to take over some or all of the tasks commonly conducted by educators? How the adaptation of guidance (the amount and the type) would affect the engagement and consequently the learning outcomes?

To this end, we built the Arigatō – an AR prototype, which aims to effectively support learners to proceed through the ExL cycle facilitated by adaptive guidance. We deliberately selected AR as the most probable future technology that will be used in the classroom since it better supports real-world in-person communication and group collaboration compared to a desktop, tablet, and VR [69, 80]. And rather than comparing the AR to other technologies our goal was to explore the design space of the AR. With the Arigatō prototype that provides options for a fixed and adaptive amount and type of guidance, we aimed to answer the following research questions:

RQ1 How do the dynamic adaptations of AR guidance influence learners' (RQ1a) recall of previously learned information, (RQ1b) mental effort, (RQ1c) task completion time, and (RQ1d) instructional efficiency?

RQ2 How does learners' engagement with AR content in terms of (i.e., task completion time, mental effort and motivation) affect their performance (i.e., recall and efficiency)?

The focus on AR in language learning covers the exploration of XR applications in the education and training domains making this study strongly relevant to the XR research community.

2 Research Background

We first identify and analyse the relevant literature on experiential learning, instructional guidance and in particular the amount and type with a focus on language learning, user engagement, and language learning in AR.

2.1 Experiential Learning

Experiential learning theory is based on the idea that learning is a process of acquiring knowledge through experience or "learning through reflection on doing" [50, 57]. Theoretical work on experiential learning continues to evolve and several models as "theories-in-use" have been proposed such as the Kolb's model [50], Boud and Walker's model [13], Joplin's five stage model [42], etc. Among these, the Kolb's experiential learning model is arguably the most influential model on educational scholarship [57]. As shown in Fig. 2, the model describes experiential learning as a four-part process, where the learner is asked to engage in a new experience, actively reflect on that experience, conceptualise it, and integrate it with prior experiences and knowledge. After completing this cycle, future decisions can be based on the newly acquired concepts.

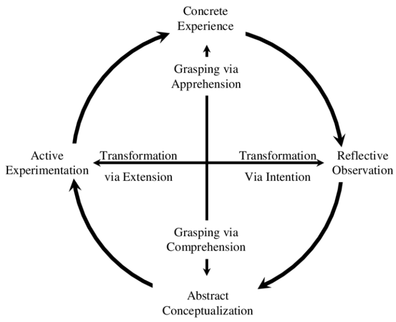


Fig. 2. Experiential learning cycle by Kolb [50]

A range of prior work has explored the use of simulated environments and AR technology to support experiential learning [3, 12, 37, 40, 67, 68, 70, 93, 97, 102]. Most of these approaches have primarily focused only on the learning stages of the Kolb's experiential learning cycle, allowing learners to perform an action and providing feedback, which learners could reflect on [12, 32, 37, 67, 68, 97]. However, according to Kolb, the learner must continue cycling through all the four stages, thus creating a "learning spiral of ever-increasing complexity" in order to gain a better learning outcome [51]. Currently, we are unaware of any work that focuses on providing guidance to support learners move from one stage to another in the experiential learning cycle and progress on the learning spiral.

2.2 Instructional Guidance

In the past, the instructional support provided during learning has been categorised and explained using different concepts and terms, such as learning paradigms (e.g. behaviourism, cognitivism and constructivism, etc.) [65, 95], instructional strategies (e.g. direct instructions, indirect instructions, interactive instructions, etc.) [66, 103] and instructional methods (e.g. lectures, case studies, peer feedback, quizzes, etc.) [22, 94]. Such paradigms, strategies and methods often emerge from different learning theories, and sometimes various belief systems and philosophies [22, 35, 64, 65, 105]. Therefore, when trying to employ a construct, such as "instructional guidance", we depend on the fact that the use of different conceptual frameworks and theories tend to define and operationalise instructional support in different ways. In the following sections, we will focus on the amount and type of guidance as described in the literature.

2.2.1 Amount of Guidance

Debates about the impact of instructional guidance and "how much" instructional support needs to be provided in a learning environment have been ongoing for at least the past century [47, 54, 64, 95]. For example, is it better to instruct learners on what they need to know by presenting them the essential information, or is it better to allow learners to discover or construct essential knowledge for themselves? Koedinger and Aleven called this issue the "assistance dilemma" [48], i.e., deciding whether to provide or withhold assistance. The contrast between the two practices can be better understood as a continuum. On one side of this continuum is the hypothesis that people learn best in an unguided or minimally guided environment [15, 90]. On the other side is the hypothesis that learners should be given instructional guidance on the concepts and procedures required for a particular task [22, 47, 64].

In a minimally guided learning environment, students are seen as active learners and are given opportunities to digest content for themselves rather than as passive learners who merely follow instructions. While these approaches have been adopted by some teachers and educators [104], there have also been decades of efforts to discourage educators from using minimally guided learning approaches. The past half-century of empirical research on the topic has provided overwhelming and unambiguous evidence that minimal guidance during learning is significantly less effective and less efficient compared to guided learning [47]. The superiority of the latter is explained in the context of human cognitive architecture and expert-novice differences [47]. Wise and O'Neill present a review of evidence to support the view that the optimal amount of guidance is often somewhere in the middle of the aforementioned continuum and that the *granularity* of the advice provided in a task (i.e. the level of details) is equally important [105].

2.2.2 Type of Instruction

Research on memory and learning has shown that comprehension and recall depend on different types of instructional methods and techniques that can be used to process and store information [25]. The mnemonic techniques have proven to be extremely effective in improving memory and recall, especially in the area of foreign language learning [6, 20, 72, 76]. Mnemonic is an instructional strategy designed for enhancing both memory and recall [62, 74, 75, 106]. This technique connects new learning to prior knowledge through the use of visual and/or acoustic cues. The basic types of mnemonic techniques rely on the use of key words, rhyming words, or acronyms [62, 75].

In the field of language learning, mnemonics have mostly been used for vocabulary learning. Despite the great variety of techniques for presenting mnemonic, the "keyword method" [11] for creating mental associations to a known language has proven to be effective in the memorisation of vocabulary. In the keyword method, learners associate the sound of a word they want to learn to one they already know in either their first language or the target language. They then mentally create an image of the known word to memorise the association [74]. This association based technique provides a powerful tool for words that have a high degree of "imagenability" [79], or for word pairs between which the learner can form some kind of semantic link [26]. The important aspect is that the keyword should clearly relate to the thing being remembered. This method also motivates learners to be more creative and use their minds more productively.

2.3 User Engagement

Engagement is a complex and multi-dimensional process [81, 83, 96] intertwined with a learners' internal indicators such as motivation, feelings, etc. (affective dimension) [30], mental effort, perceptions, etc. (cognitive dimension) [30], and observable actions such as performing various activities, interacting with a system, etc. (behavioural dimension) [8, 30]. In addition, Reeve and Tseng suggested incorporating agentic engagement as a fourth dimension of engagement [77]. Agentic engagement refers to the proactive and intentional activity of the learner to personalise the conditions of learning and to enrich external learning goals.

Previous studies have shown that behavioural, cognitive and affective dimensions predict learner' performance both separately and in unison [4, 96]. This means that the dimensions of user engagement are interrelated and simultaneously affect human behaviour [30].

Kearsley and Shneiderman have stressed that engagement can be stimulated without technology, but digital technology opens up novel possibilities that are hard to achieve in a physical form [46]. Accordingly, research in this area has explored engagement with images [29], video clips [2, 89], music [39, 49], and real-life scenarios [45, 99]. However, while there are a limited numbers of studies such as [33, 96], there is a lack of research on learning and engagement in AR environments.

2.4 Language Learning in AR

A considerable body of literature focuses on guidance in AR environments to support learning in general [12, 60, 63, 93, 101] and language learning in particular [9, 23, 24, 38, 87, 107]. Prior work on language Learning in AR is summed in Table 1. We categorised it based on the following dimensions: (i) *Hardware used:* mobile AR, HMD, sensors; *Learning focus:* vocabulary and/or grammar; *Guidance method:* generic or adaptive; and *Learning method:* experiential, contextual, game-based learning, etc. [100]). Importantly, most of the existing approaches deliver generic guidance, in which all users receive the same set of general instructional cues, and in the same way [12, 60, 63, 101].

Table 1. Selected prior work related to language learning in AR environments and how our work differs alongside different dimensions.

Study	Hardware used	Learning Focus	Guidance Method	Learning Method
Draxler et al. (2020) [24]	Mobile	Grammar	Generic	Context-based learning
Arvanitis et al. (2020) [9]	Mobile	Vocabulary	Generic	Self-directed learning
Yang & Mei (2018) [107]	Mobile	Vocabulary	Generic	Game-based learning
Ibrahim et al. (2018) [38]	HMD	Vocabulary	Generic	Context-based learning
Vazquez et al. (2017) [98]	HMD	Vocabulary	Generic	Context-based learning
Dita (2016) [23]	Mobile	Vocabulary	Generic	Game-based learning
Seedhouse et al. (2014) [87]	Sensors	Vocabulary	Generic	Experiential learning
Liu & Tsai (2013) [59]	Mobile	Vocabulary	Generic	Context-based learning
Arigatō (2022)	HMD	Vocabulary Grammar	Adaptive	Experiential learning

However, as pointed out in Sect. 2.2, the amount and the type of guidance play a vital role when providing instructional guidance. In addition, the adaptation of guidance to user needs is also important as noted in Sect. 2.2.1. Together they affect the engagement and consequently the learning outcomes of each individual learner. While there are studies such as [36, 93] that provide adaptive guidance for performing a physical task, generally there is a lack of research on providing adaptive guidance based on the amount or type of instructions. Moreover, we are unaware of any work that focuses on the "keyword method" as the type of guidance used to support language learning in AR environments.

In order to address these gaps, we developed an adaptive guidance AR system for language learning called Arigatō. We employ this prototype to investigate the effect of the amount (fixed vs. adaptive) and the type of associations (fixed vs. adaptive) on the engagement and consequently the learning outcomes of language learning in an AR environment. The research method to investigate the aforementioned effects is presented in the next section.

3 Research Method

The Arigatō prototype for language learning was developed to answer our research questions. In this section, we present the prototype and we describe the study conditions, study design, study procedure, participants' sampling, data collection, and analysis.

3.1 Study Conditions

The language selected for this study was Japanese, since it is unrelated to the Indo-European family of languages and we expected people would not be familiar with its grammar and vocabulary. We designed four different study conditions based on two different aspects of instructional guidance that can be adapted: the AMOUNT of guidance (FIXED-AMOUNT, which is the same for all, and ADAPTIVE-AMOUNT, which decreases based on the learners' performance), and the TYPE of instructions, (FIXED-ASSOCIATIONS using predefined 3D AR models of the vocabulary being learnt and ADAPTIVE-ASSOCIATIONS using self selected 3D AR models to create associations ("keyword method") to the vocabulary being learnt).

The structure of the design of the study with all four (4) study conditions is illustrated in Fig. 3. The study was planned as a 2 x 2 mixed design study (including both within and between-subjects) envisaged to take approximately 60 to 75 minutes. A common within-subjects

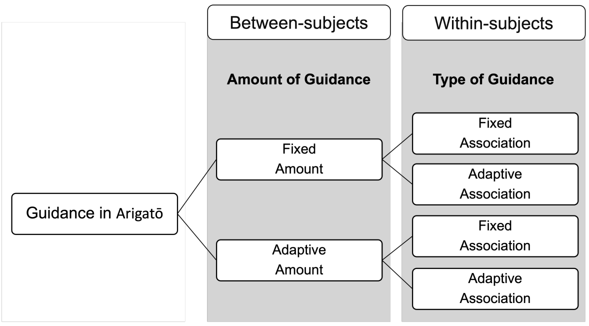


Fig. 3. Study design and conditions.

design would make this cognitively demanding learning study even longer (approx two to three hours), which might hinder participants' performance and negatively affect the results. Other options such as splitting the study into several sessions would also introduce other biases (e.g. users might study between sessions, day to day performance might vary) and practical issues (e.g. getting all users back for the following session).

The AMOUNT of guidance was thus evaluated as a within-subjects variable while the TYPE of instructions as a between-subjects variable. This means that each participant either received the FIXED-AMOUNT or the ADAPTIVE-AMOUNT of guidance, but all participants experienced both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS.

In each condition, participants had to learn Japanese vocabulary and grammar around a particular topic (Christmas or birthday celebration) by learning and understanding four (4) distinct phrases and their structure, and recall them successfully. The phrases related to Christmas were: family gathering, Christmas tree preparation, turkey dinner and lights decorations. The phrases related to birthday celebration were: invite friends, blow balloons, bake a cake and light up candles.

In the FIXED-AMOUNT condition, at each stage of the learning cycle participants received all the instruction for all the content needed to be learnt for all four (4) phrases repeatedly through consequent cycles until all phrases were correctly recalled. In the ADAPTIVE-AMOUNT condition, at each stage of the learning cycle participants only received instructions for the phrases that were not recalled correctly in previous cycles. Thus, once a phrase was recalled correctly in the CE stage, the guidance for that phrase was not shown in the next cycles. In the FIXED-ASSOCIATIONS condition, participants were presented with predefined 3D AR models of objects corresponding to vocabulary of the phrases being learned. In the ADAPTIVE-ASSOCIATIONS condition, the participants could self-select 3D AR models to create associations for the corresponding vocabulary of the phrases being learned.

To avoid the "order effects" (the influence of the order in which the conditions are presented on participants' performances [86]), the order of TYPE of instructions (FIXED and ADAPTIVE) as well as the order of the topic being learnt (Christmas and birthday celebration) was balanced among the participants.

3.2 The Arigatō Prototype

The prototype was developed for the Magic Leap one AR head mounted display (HMD) with controller inputs [1], using the Unity3D game development environment [2]. The Lumin SDK 0.26 and Magic leap tools were used for setting up the development environment with Lumin OS 0.98.3. The MRTK Mixed Reality tool kit [3] was used for integrating the inputs and object manipulation techniques. Speech recognition was implemented with the IBM Watson speech to text API [4].

[1]https://www.magicleap.com/
[2]https://unity.com/
[3]https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity
[4]https://cloud.ibm.com/apidocs/speech-to-text

The prototype replicates all four stages of Kolbs' experiential cycle. Before entering the cycle, the learner first sees an introduction window including all four (4) phrases related to the topic being learnt in English in order to familiarise themselves with the intent and content of the study. Besides text instructions and explanations, audio instructions and explanations are also provided throughout the cycle.

Fig. 4. The study setup where the participant tries to finish the learning tasks visible in AR HMD and researcher follows the process on the laptop screen.

The system starts the learning cycle in the abstract conceptualisation (AC) stage. The AC stage provides the description of the vocabulary and the basic grammar rules related to phrases being learnt. The vocabulary is presented by either (i) FIXED-ASSOCIATIONS (for example, for the word lights – "raito" in Japanese – a 3D virtual object of a real world light set is shown by the word as visible in Fig. 1b), or (ii) ADAPTIVE-ASSOCIATIONS (for example, for the word candle – "rosoku" in Japanese – a merged 3D virtual object/a combined metaphor of a rose inside a sock is shown by the word as visible in Fig. 1f). After familiarising themselves with the content in the AC stage, the learner proceeds to the active experimentation (AE) stage.

In the AE stage, the system shows each English phrase with a puzzle task to generate the corresponding Japanese phrase made of three parts from six (6) possibilities offered. The puzzle is also presented in two ways based on the TYPE of instructions: (i) a word puzzle with text components only (Fig. 1c) or (ii) a word puzzle together with 3D virtual objects for the learner to create associations (Fig. 1e) from a given set of objects. If the learner does not know the solution to the puzzle they can skip that phrase. Feedback for the correct and incorrect phrases is given as text and highlights as shown in Fig. 1c and Fig. 1e. There, if the answer is incorrect, it highlights the given answer in red and shows the correct answer separately.

After completing the puzzles related to all four phrases, the system nudges the learner to proceed to the concrete experience stage (CE). A 3D avatar is shown asking the learner to recall each phrase in Japanese (Fig. 1d) by speaking it aloud. To evaluate the spoken answers, the IBM Watson speech recognition was not accurate enough, so a Wizard of Oz technique was used for the experiment. If the phrase is correctly voiced, the corresponding 3D AR model is displayed. The learner can manipulate the object and place it anywhere in the room as they would be able to do it in a real-world experiential learning scenario with physical objects.

If all the phrases are not recalled correctly in the CE stage, the system automatically moves to the reflective observation (RO) stage. There, the learner can first go through all the phrases that include corresponding visual models and audio pronunciation, and then recall them by selecting the correct Japanese phrase (out of four phrases) corresponding to the given English phrase (Fig. 1a and g). The system then returns back to the AC stage.

In this way, the system nudges the learner to proceed further in the experiential learning cycle and consequently on the learning spiral until all four (4) phrases are recalled correctly in the CE stage. In the FIXED-AMOUNT condition, all learners have to repeat all phrases in all stages in each cycle in the same way. In the ADAPTIVE-AMOUNT condition, the learner has to go only through and repeat previously incorrectly recalled phrases (in the CE stage) as the content related to previously correctly recalled phrases is no longer shown.

One of the main features of the Arigatō prototype is the possibility to move digital objects around the physical space as well as move various elements of the interface (e.g. puzzle) in order to complete the tasks given. It is this manipulation of the digital elements that can be brought in the visual field and removed that supports experiential learning or learning by doing.

3.3 Participants

In total, 28 participants were recruited for the study by invitation in a mailing list and web page as well as by snowball sampling. All were students and staff members from our university whose first language is one of the Indo-European languages. None of the participants had any prior knowledge in the Japanese language (identified via a short competency test questionnaire). The between-subject experimental sample comprised 14 participants in the FIXED-AMOUNT condition with 6 (46%) females and 14 participants in the ADAPTIVE-AMOUNT condition with 8 (54%) females. All participants were between 18 to 31 years old ($\bar{x} = 25$) and randomly assigned to one of the two conditions. The study was approved by the local Research Ethics Board.

3.4 Procedure

Participants were first given a consent form to sign, together with the Participant Information Sheet (PIS) outlining the entire research process, and were given an opportunity to ask any question related to the study. They were also instructed that they could abandon the study at any stage. Next, they were requested to fill out the Questionnaire on Current Motivation (QCM).

Before starting the actual task, they completed a five minute training session on a demo application to understand the interface and the interaction with the system. Participants were then instructed to complete a language learning task on one topic (Christmas or birthday celebration) in one condition (either FIXED-ASSOCIATIONS or ADAPTIVE-ASSOCIATIONS) and the other topic in the other condition (see within-subject part of the Fig. 3). After finishing each topic, participants filled out a mental effort questionnaire and a recall questionnaire (to assess their immediate recall) asking participants to remember phrases they just tried learn. They were given a 5 minutes break in between the topics.

In addition, at the end of the study, participants answered two standard questionnaires: a system usability (SUS) [56] and a user experience questionnaire (UEQ) [85]. They also filled out a short post-questionnaire with demographic questions, questions about previous experience with AR technology, and questions about their vision. The whole experiment lasted from 60 to 75 minutes.

One week after the study, participants were again requested to answer the same recall questionnaire (as just after finishing the study) to assess their delayed recall.

3.5 Data Collection

In all conditions, the performance progression data and the time stamp data were logged by the system as a part of behaviour engagement. To measure the motivation, the short form of the Questionnaire on Current Motivation (QCM) with 12 items/questions [31, 78] was used. QCM measures anxiety, challenge, interest, and probability of success on a five-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Instead of focusing in individuals sub-dimensions (i.e., anxiety, challenge, interest, and probability of success), we used the mean score of the 12 items as an indicator of the overall motivation.

For measuring mental effort we used a standard 8-questions questionnaire from [43] that focuses on mental load for learning contexts. To ensure comparability with other studies while maintaining coherence with our tasks, we just changed the wording to reflect the task at hand and better fit our learning scenario. In addition, we added 5 questions relevant to our study (How much effort did you invest in (i) memorising the words, (ii) understanding particles, (iii) recalling the memorised words, (iv) building the sentences with puzzle blocks, and in (v) speaking the sentences out loud?). All 13 questions were answered on a nine-point scale. The mean scores of the 8 (MEQ8) and the 5 (MEQ5) answers as well as the overall mean score (MEQ13) were used as indicators for the mental effort.

The recall questionnaire (immediate and delayed) included Japanese phrases the participants practised during the learning tasks (e.g., "How do you say 'invite friends' in Japanese?", "How do you say 'family gathering' in Japanese?" with the result marked as either correct or incorrect).

To measure the usability of the system, we used the System Usability Scale (SUS) questionnaire [14] (10 questions answered on a five-point Likert scale). The SUS scores were calculated with the standard SUS analysis. For measuring the user experience, we used the short version of the User Experience Questionnaire (UEQ-S) [84, 85] with eight items/questions. The first four represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four hedonic qualities (Stimulation and Novelty) [84]. The results are converted to a range between -3 to 3.

To assess the reliability of motivation and mental effort questionnaires, we performed Cronbach's alpha test. Estimated reliability for each questionnaire (motivation Cronbach's $\alpha = 0.71$ and mental effort $\alpha = 0.89$) was acceptable for the research purposes [10]. For the recall questionnaire, the reliability was measured using the Kuder-Richardson 20 test [53] because of the binary nature of the results (correct/incorrect). The $KR = 0.76 > 0.5$ value indicates that the reliability of the recall questionnaire was also acceptable.

3.6 Data Analysis

Each data set collected in the study was first checked for normality using the Shapiro–Wilk normality test [88].

The analysis was done in R studio, using "WRS2" R package. For immediate recall, delayed recall, mental effort, task completion time and learning efficiency, the statistical significance was examined using a mixed between-within subjects ANOVA on the 20% trimmed means–"bwtrim" [61]. The main between-subjects effect (group comparisons), the main within-subjects effect (e.g., due to repeated measurements), and the interaction effect, were computed using the "sppba", "sppbb", and "sppbi" functions respectively [61]. Statistical significance for motivation was examined using the Mann-Whitney U test [91].

The resulting $p < 0.05$ are reported as statistically significant. All boxplots use a 1.5xIQR (interquartile range) rule and Tukey's fences [92] for whiskers and identified outliers. Asterisk notation is used in figures to visualise statistical significance (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$ and ***: $p < 0.001$).

We also conducted a power analysis to check and validate the results and findings of the study. We calculated the effect size (Cohen's d) for each data set collected [21], selected the minimum effect size (Cohen's $d = 1.251$) and estimated the statistical power ($1 - \beta = 0.8$) of data to check whether the type II error probability (β) is within an acceptable range for a given sample size ($n = 14$ per group) and a significance level ($\alpha = 0.05$). The estimated power value 0.8 shows that with the given sample size, we can have a 80% chance that we correctly reject the null hypothesis with a significance level 0.05.

Finally, we used a Pearson's multiple correlation test [73] to find out whether there were any correlations between a learner's mental effort, task completion time, motivation and their performances (i.e., recall and efficiency).

4 Results

The results of the analysis based on the input from the 28 participants who completed the study is presented here. Before dwelling further into the effects of guidance on various measures we have checked whether the motivation has had an effect on the study. The results show that the motivation had no statistically significant effect on the rest of the results presented in this section (Mann-Witney U test, $U(N_{\text{FIXED-AMOUNT}} = 14, N_{\text{ADAPTIVE-AMOUNT}} = 14) = 90.20$, $p > 0.05$). The following sections present the effect of adaptive guidance (amount and type) on learning performance (immediate and delayed recall), mental effort, task completion time, and learning efficiency (immediate and delayed). The last section presents correlations between these measures.

4.1 The Effect of Guidance on Learning Performance

Learning performance was measured with the recall questionnaires right after the study (immediate) and after 7 days (delayed). The results are presented in the following two sections.

4.1.1 Immediate recall

The effect of the adaptive guidance – i.e., the AMOUNT (FIXED and ADAPTIVE) and TYPE (FIXED and ADAPTIVE) – on immediate recall is shown in Fig. 5 (top left). The data summarised in Fig. 5 (bottom left) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results in Fig. 5 (bottom left) show that the effects of the AMOUNT and TYPE of guidance on immediate recall are statistically significant. The effect of the AMOUNT of guidance alone ($df = 14.19$, $p < 0.001$) is significant. Fig. 5 (top left) indicates that the immediate recall in the FIXED-AMOUNT condition ($\bar{x} = 83.04\%$, $SD = 16.44$) is significantly better compared to the ADAPTIVE-AMOUNT condition ($\bar{x} = 66.7\%$, $SD = 19.15$).

Similarly, the effect of the TYPE on immediate recall ($df = 17.88$, $p < 0.001$) is also significant. Furthermore, Fig. 5 (top left) indicates that the immediate recall in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 86.61\%$, $SD = 11.57$) is significantly better compared to the FIXED-ASSOCIATIONS condition ($\bar{x} = 62.5\%$, $SD = 14.42$). The interaction effect of the AMOUNT and TYPE on the immediate recall ($df = 17.88$, $p > 0.05$) is not statistically significant.

4.1.2 Delayed recall

The effect of the adaptive guidance (AMOUNT and TYPE) on delayed recall is shown in Fig. 5 (top centre). The data summarised in Fig. 5 (bottom centre) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results show that the effects of the AMOUNT and TYPE of guidance on delayed recall are statistically significant. The effect of the AMOUNT of guidance alone ($df = 15.13$, $p < 0.01$) is also significant. Fig. 5 (top centre) indicates that the delayed recall in the FIXED-AMOUNT condition ($\bar{x} = 70.54\%$, $SD = 23.20$) is significantly better compared to ADAPTIVE-AMOUNT condition ($\bar{x} = 50.00\%$, $SD = 23.15$).

The effect of the TYPE on the delayed recall ($df = 17.68$, $p < 0.001$) is also significant. Fig. 5 (top centre) indicates that the delayed recall in the ADAPTIVE-ASSOCIATION condition ($\bar{x} = 73.21\%$, $SD = 18.12$) is significantly better compared to FIXED-ASSOCIATION condition ($\bar{x} = 47.32\%$, $SD = 20.24$). The interaction effect of the AMOUNT and TYPE on the delayed recall ($df = 17.68$, $p > 0.05$) is not significant.

4.2 The Effect of Guidance on Mental Effort

The effect of the adaptive guidance (AMOUNT and TYPE) on mental effort is shown in Fig. 5 (top right). The data summarised in Fig. 5 (bottom right) is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means [61].

The results show that the effects of the AMOUNT and TYPE of guidance on mental effort are statistically significant. The effect of the AMOUNT of guidance alone (MEQ13 $df = 17.47$, $p < 0.01$; MEQ08 $df = 16.09$, $p < 0.01$; MEQ05 $df = 14.73$, $p < 0.01$) is significant. Fig. 5 (top right) indicates that the overall mental effort in the ADAPTIVE-AMOUNT condition ($\bar{x} = 3.66\%$, $SD = 1.06$) is significantly lower compared to the FIXED-AMOUNT condition ($\bar{x} = 4.61$, $SD = 0.98$).

The effect of the TYPE alone is also significant (MEQ13 $df = 17.78$, $p < 0.001$; MEQ08 $df = 12.58$, $p < 0.01$; MEQ05 $df = 17.73$, $p < 0.01$). Fig. 5 (top right) indicates that the overall mental effort in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 3.39$, $SD = 0.63$) is significantly lower compared to the FIXED-ASSOCIATIONS condition ($\bar{x} = 4.88$, $SD = 0.73$). The interaction effect of the AMOUNT and TYPE on mental effort (MEQ13 $df = 17.78$, $p > 0.05$; MEQ08 $df = 12.58$, $p > 0.05$; MEQ05 $df = 17.73$, $p > 0.05$) is not statistically significant.

4.3 The Effect of Guidance on Task Completion Time

The results for the task completion time presented in Fig. 6 (bottom left) are analysed using a mixed between-within subjects ANOVA on the 20% trimmed means. The results indicate that the AMOUNT of guidance has no statistically significant effect on the task completion time ($df = 15.29$, $p > 0.05$). It can thus not be concluded that participants in the ADAPTIVE-AMOUNT condition completed the task faster compared to participants in the FIXED-AMOUNT condition.

In contrast, the effect of the TYPE on the task completion time is significant ($df = 17.88$, $p < 0.001$) as can be also observed in Fig. 6 (top left). This also shows that the task completion time in the ADAPTIVE-ASSOCIATIONS condition ($\bar{x} = 15.78$ *min*, $SD = 3.17$) is significantly lower compared to the FIXED-ASSOCIATION condition ($\bar{x} = 27.13$ *min*, $SD = 6.69$). The interaction effect of the AMOUNT and TYPE on the task completion time is not statistically significant ($df = 17.88$, $p > 0.05$).

4.4 The Effect of Guidance on Learning Efficiency

The learning efficiency for each study condition was determined using Equation 1 [18, 34, 71].

$$E = \frac{z_P - z_M}{\sqrt{2}} \quad (1)$$

E = Learning efficiency
z_P = Average performance in Z-scores
z_M = Average task difficulty in Z-scores

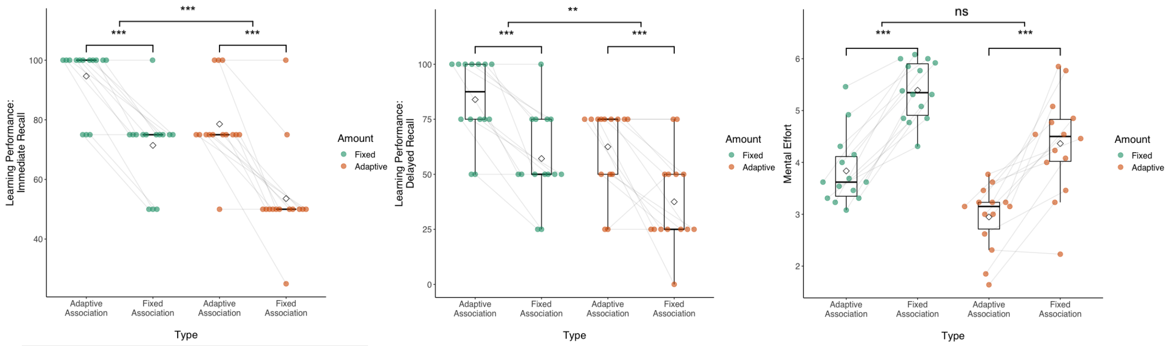
We measured the performance of each study condition based on the recall scores participants obtained after completing the task for that study condition. We estimated the difficulty of the task based on the mental effort questionnaire for that study condition. We then calculated the learning efficiency for each of the four conditions: (a) FIXED-AMOUNT of guidance with FIXED-ASSOCIATIONS, (b) FIXED-AMOUNT of guidance with ADAPTIVE-ASSOCIATIONS, (c) ADAPTIVE-AMOUNT of guidance with FIXED-ASSOCIATIONS, and (d) ADAPTIVE-AMOUNT of guidance with ADAPTIVE-ASSOCIATIONS using Formula 1. This was done for both immediate recall (immediately after participants had completed the task) and delayed recall (a week after participants had completed the task).

4.4.1 Immediate Efficiency

The immediate recall learning efficiency results across the study conditions are depicted in Fig. 7 (left). It shows that, for both FIXED-AMOUNT and ADAPTIVE-AMOUNT, the ADAPTIVE-ASSOCIATIONS condition is more efficient (i.e., has a higher efficiency value) compared to the FIXED-ASSOCIATIONS condition.

The effects of the AMOUNT and TYPE of guidance, on immediate recall learning efficiency are shown in Fig. 6 (top centre). The data is analysed using a mixed between-within subjects ANOVA on the 20% trimmed means (Fig. 6 (bottom centre)). There is no significant effect of the AMOUNT of guidance conditions on participants' immediate recall learning efficiency ($df = 16.80$, $p = 0.460$).

By contrast, the effect of the TYPE of instructions on immediate recall learning efficiency is statistically significant ($df = 16.80$, $p < 0.001$). Fig. 6 (top centre) indicates that the immediate recall learning efficiency in the ADAPTIVE-ASSOCIATIONS condition is significantly higher compared to the FIXED-ASSOCIATIONS condition. The interaction effect of the AMOUNT and TYPE on the immediate recall learning efficiency is not statistically significant ($df = 16.80$, $p > 0.05$).

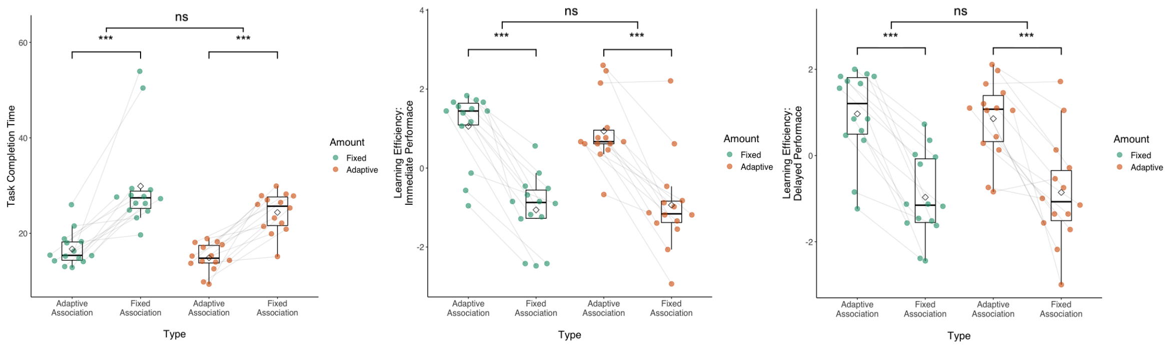


	Immediate recall			
	df_1	df_2	F	p
Amount	1	14.19	26.58	0.000
Type	1	17.88	91.09	0.000
Amount : Type	1	17.88	0.21	0.655

	Delayed recall			
	df_1	df_2	F	p
Amount	1	15.13	12.60	0.002
Type	1	17.68	92.65	0.000
Amount : Type	1	17.68	0.15	0.706

	Mental Effort			
	df_1	df_2	F	p
Amount	1	17.47	13.32	0.002
Type	1	17.78	171.13	0.000
Amount : Type	1	17.78	2.42	0.137

Fig. 5. Left: Immediate recall in percentage of correctly remembered phrases. Centre: Delayed recall in percentage of correctly remembered phrases. Right: Overall mental effort (MEQ13) invested during the task. Below are results form mixed between-within subjects ANOVA on the 20% trimmed means examining the between-subjects (AMOUNT), the within-subjects (TYPE) and, the interaction (AMOUNT:TYPE) effects, over dependent variables.



	Task Completion Time			
	df_1	df_2	F	p
Amount	1	15.29	2.62	0.126
Type	1	17.88	189.77	0.000
Amount : Type	1	17.88	0.94	0.346

	Learning Efficiency: immediate recall			
	df_1	df_2	F	p
Amount	1	16.69	0.66	0.427
Type	1	16.80	183.92	0.000
Amount : Type	1	16.80	1.53	0.233

	Learning Efficiency: delayed recall			
	df_1	df_2	F	p
Amount	1	17.74	0.10	0.756
Type	1	16.72	151.92	0.000
Amount : Type	1	16.72	0.57	0.460

Fig. 6. Left: Task completion time in minutes. Centre: Average learning efficiency for immediate performance. Right: Average learning efficiency for delayed performance. Below are results form mixed between-within subjects ANOVA on the 20% trimmed means examining the between-subjects (AMOUNT), the within-subjects (TYPE) and, the interaction (AMOUNT:TYPE) effects, over dependent variables.

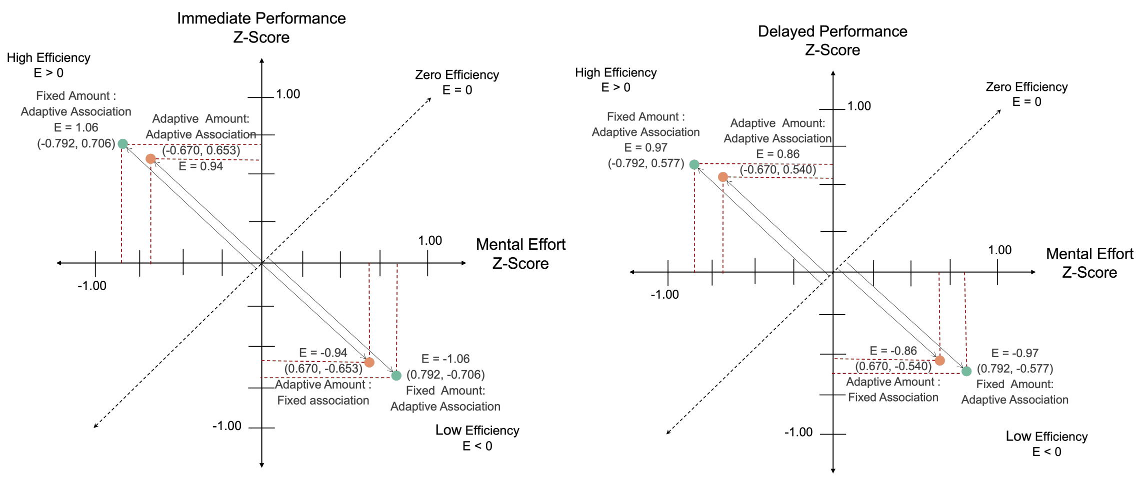


Fig. 7. Left: Immediate recall learning efficiency. Right: Delayed recall learning efficiency.

4.4.2 Delayed Efficiency

The delayed recall learning efficiency results across the study conditions are shown in Fig. 7 (right). The figure indicates that, for both FIXED-AMOUNT and ADAPTIVE-AMOUNT, the ADAPTIVE-ASSOCIATION is more efficient (i.e., has a higher efficiency value) compared to the FIXED-ASSOCIATIONS condition.

The effects of the AMOUNT and TYPE of guidance on delayed recall learning efficiency are illustrated in Fig. 6 (top right).

There is no significant effect of the AMOUNT of guidance condition on the delayed recall learning efficiency ($df = 16.72$, $p = 0.460$).

By contrast, the effect of the TYPE of instructions on the delayed recall learning efficiency is statistically significant ($df = 16.72$, $p < 0.001$). The results also show that the delayed recall learning efficiency is significantly higher for ADAPTIVE-ASSOCIATIONS compared to FIXED-ASSOCIATIONS condition. No statistically significant interaction effect was found between conditions ($df = 16.72$, $p > 0.05$).

4.5 System Usability and User Experience

The average SUS score shows that the system usability is in an acceptable range (FIXED-AMOUNT = 82.3, ADAPTIVE-AMOUNT = 85.8; $SUS > 70$). The results for UEQ show that pragmatic (FIXED-AMOUNT = 1.54, ADAPTIVE-AMOUNT = 1.75) and hedonic qualities (FIXED-AMOUNT = 2.33, ADAPTIVE-AMOUNT = 2.52), as well as overall user experience (FIXED-AMOUNT = 1.93, ADAPTIVE-AMOUNT = 2.14) are perceived as strongly positive in both conditions. In all cases we see that the adaptive-amount has higher values.

4.6 Correlations between Measures

A Pearson multiple correlation test was applied to find out whether or not there were any correlations between the measures (recall, efficiency, mental effort, task completion time). The results are presented in Table 2.

The results for the FIXED-AMOUNT condition indicate that in both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS conditions, the mental effort is negatively related to both immediate and delayed recall as well as to both immediate and delayed efficiency. The task completion time measure does not show any relationship to other variables. The correlations between immediate/delayed and recall/efficiency variables was expected since these measures are all related.

Similar results can be observed for ADAPTIVE-AMOUNT condition. In both FIXED-ASSOCIATIONS and ADAPTIVE-ASSOCIATIONS conditions, the mental effort is negatively related to both immediate and delayed recall as well as to both immediate and delayed efficiency. The task completion time does also not show any relationship to other variables. While, as before, immediate/delayed and recall/efficiency correlations are not surprising.

5 DISCUSSION

Arigatō was developed to compare how learners perform in learning a new language, when presented with different adaptation factors of AR based instructional guidance in an experiential learning environment. The AR system was deliberately selected as the most probable future technology that will be used in the classroom. An AR classroom has several advantages over other types of digitised classrooms such as taking the real world into consideration and consequently embedding information directly into the user's field of view. These advantages better support real-world in-person communication and group collaboration compared to other technologies AR is often contrasted to, such as desktop, tablet computers, and VR [69, 80].

By taking the real world into consideration, our prototype did not just show content on the HMD, but allowed users to move and place AR objects in their physical surroundings. With this functionality, AR was used to replicate the real-world experiential learning scenario (e.g., learning by decorating a room for Christmas celebrations with AR objects as can be done with real objects) and show information in a coherent and meaningful way within the real world context.

Arigatō allowed us to vary (i) the AMOUNT of guidance, which is either FIXED or an ADAPTIVE-AMOUNT (the amount changes based on participants' performance), and (ii) the TYPE of association, which is either FIXED or with ADAPTIVE-ASSOCIATIONS (participants can select associations by themselves from a predefined set). We used a 2 × 2 mixed design to evaluate the study conditions, as explained in Sect. 3.1. The AMOUNT of guidance was evaluated as a within-subjects variable while the TYPE of instructions as a between-subjects variable. In this paper, the impact of the guidance method was measured through performance (i.e., recall and efficiency) and engagement (i.e., task completion time, mental effort, and motivation).

We measured behavioural engagement for FIXED guidance and ADAPTIVE guidance by observing how participants interacted with Arigatō and how much time they spent in each condition. We could observe participants being more interested in the ADAPTIVE-ASSOCIATIONS than the FIXED-ASSOCIATIONS condition, even though they completed the task in less time. This was expected, since keyword methods have been suggested as motivating learners to be more creative and to enjoy using their minds more productively [79].

We measured cognitive engagement for all the study conditions by analysing how much mental effort participants invested in completing the task. We found that the mental effort with the ADAPTIVE-AMOUNT of guidance was significantly lower compared to the FIXED-AMOUNT of guidance. This is in line with the literature, as prior work has shown that repeating the same information could increase the amount of cognitive work and consequently the mental effort [82]. We further identified that the mental effort in the ADAPTIVE-ASSOCIATIONS condition was lower than in FIXED-ASSOCIATIONS condition. Again, this was somewhat expected, as the association method connects new learning to prior knowledge through the use of visual and/or acoustic cues [62, 75], which could lower the mental effort. However, we have shown that in the context of language learning the AR environment can support the creation of adaptive constructs not only in the form of the keywords (the "keyword method" [11]) and mental associations, but also in combination with 3D digital objects that users can manipulate with. It is this combination that supports experiential learning through manipulation of digital objects and that outperformed the traditional "keyword method" in our study.

We measured learning performance through analysis of recall (immediate recall after the study and delayed recall test 7 days after the study) and learning efficiency. Interestingly, we found that compared to the ADAPTIVE-AMOUNT of guidance, the FIXED-AMOUNT of guidance group scored better in the immediate and delayed recall tests, although they have invested a significantly higher mental effort during the task. In addition, we found that the ADAPTIVE-ASSOCIATIONS outperformed the FIXED-ASSOCIATIONS group, in the immediate, delayed recall tests, and learning efficiency, while investing a significantly lower mental effort and spending a less amount of time during the task. These findings suggest avenues for future work in archetypal user models, which might allow users to slowly adapt pre-developed personal models from a standard set of ideal users. What we would refer to as the "follow the leader" learning approach.

We have also investigated how different variables are correlated and how they could be combined in order to more accurately present engagement and learning performances of the participants' for each study condition. This contrasts to the existing literature in learning that often looks at these aspects individually. Our results show correlations between *mental effort and learning performances (both recall and efficiency)* in all conditions, which was to be expected since the two are considered related. Interestingly, there were no correlations found between *task completion time and learning performance* in any condition. Again, these findings suggest potential future work in "transfer motivation" between individual learners or between instructors and learners.

6 LIMITATIONS AND FUTURE WORK

The average age of participants in our study was 25 years, which presents a possible age bias. Despite this, this age group is worth studying as it is highly mobile, spending an extended period of time in a foreign speaking country (for example, the EU Erasmus+ programme alone funds more than half a million exchanges yearly [27]). As such,

Table 2. Correlations between different measures for the FIXED-AMOUNT of guidance condition (left) and the ADAPTIVE-AMOUNT of guidance condition (right). ME: mental effort, CT: task completion time, IR: immediate recall, DR: delayed recall, IE: immediate efficiency and DE: delayed efficiency.

FIXED-AMOUNT of guidance							ADAPTIVE-AMOUNT of guidance						
	FIXED-ASSOCIATIONS							FIXED-ASSOCIATIONS					
	ME	CT	IR	DR	IE	DE		ME	CT	IR	DR	IE	DE
ME	1						ME	1					
CT	-0.233^{ns}	1					CT	-0.546^{ns}	1				
IR	−0.627	-0.014^{ns}	1				IR	−0.820	0.278^{ns}	1			
DR	−0.819	0.046^{ns}	0.797	1			DR	−0.813	0.393^{ns}	0.679	1		
IE	−0.863	0.097^{ns}	0.935	0.891	1		IE	−0.956	0.436^{ns}	0.951	0.784	1	
DE	−0.931	0.125^{ns}	0.764	0.972	0.920	1	DE	−0.952	0.493^{ns}	0.786	0.953	0.913	1
	ADAPTIVE-ASSOCIATIONS							ADAPTIVE-ASSOCIATIONS					
	ME	CT	IR	DR	IE	DE		ME	CT	IR	DR	IE	DE
ME	1						ME	1					
CT	0.154^{ns}	1					CT	-0.208^{ns}	1				
IR	−0.846	-0.075^{ns}	1				IR	−0.884	0.135^{ns}	1			
DR	−0.903	-0.338^{ns}	0.745	1			DR	−0.618	0.432^{ns}	0.568	1		
IE	−0.963	-0.121^{ns}	0.958	0.860	1		IE	−0.966	0.174^{ns}	0.975	0.609	1	
DE	−0.972	-0.260^{ns}	0.812	0.979	0.931	1	DE	−0.863	0.375^{ns}	0.777	0.930	0.841	1

*** Correlation is significant at 0.001 level, ($p < 0.001$)
** Correlation is significant at 0.01 level, ($p < 0.01$)
* Correlation is significant at 0.05 level, ($p < 0.05$)
ns Correlation is not significant, ($p > 0.05$)

*** Correlation is significant at 0.001 level, ($p < 0.001$)
** Correlation is significant at 0.01 level, ($p < 0.01$)
* Correlation is significant at 0.05 level, ($p < 0.05$)
ns Correlation is not significant, ($p > 0.05$)

this group could benefit from an improved language learning system. Nevertheless, the results cannot be generalised over the whole population, and expanding the study to other age groups and exploring the effect of age on the proposed learning system is an important future direction.

Another bias is the self-selection bias because the participants who opted-in are likely to be at ease when learning a new language. However, even if this study attracted only a particular group of language learners, the results still show the benefit of AR guidance with adaptive associations that would most likely benefit also other users in our studied age group. However, this needs to be further investigated. Using questionnaires for measuring certain aspects such as motivation and mental effort can result in social desirability bias as users could answer them to please the researcher. However, we used well established questionnaires together with measurements that can hardly be affected by such bias (i.e., recall and efficiency and task completion time).

The Arigatō prototype covers the first three levels (categories) of the Bloom's taxonomy [1] only: Knowledge (remembering the words, sentences), Comprehension (understanding the structure of sentences), and Application (applying or speaking out the sentences). The expansion to higher levels of the taxonomy (Analysis (analyse), Synthesis, and Evaluation (evaluate), Creation) is also relevant and we aim to take this into consideration in future implementations. In addition, we used the task of learning vocabulary and other language constructs as a use-case to explore the potential of AR in adaptive learning scenarios. In future, this use-case should be expanded to other use-cases to confirm its generalisability.

7 Conclusion

In this paper we present the Arigatō (Augmented Reality Instructional Guidance & Tailored Omniverse) prototype – an adaptive guidance augmented reality (AR) system for language learning. AR has been identified as an ideal platform to support simulating various experiences for experiential learning and we explored the AR's design space in the learning context rather than comparing it to other technologies. With our prototype we investigated how the amount of guidance (fixed vs. adaptive-amount) and the type of guidance (fixed vs. adaptive-associations) affects the engagement and consequently the learning outcomes of language learning in an AR environment.

Compared to the adaptive-amount, the fixed-amount of guidance group scored better in the immediate and delayed (after 7 days) recall tests. However, this group also invested a significantly higher mental effort to complete the task. Adaptive-associations group outperformed the fixed-associations group in the immediate, delayed (after 7 days) recall tests, and learning efficiency. The adaptive-associations group also showed significantly lower mental effort and spent less time to complete the task. Both results hint at potential for the archetypal user models in comparable learning scenarios, which might allow users to "transfer motivation" between individual learners or between instructors and learners, and slowly adapt pre-developed personal models. Such approach could achieve the balance between the (adaptive) amount and (adaptive) type to optimise the mental effort need to complete the learning task.

The results also show the potential of AR in adaptive learning scenarios where the learning environment needs to be simulated and where virtual content needs to be added or removed on the fly based on the learning needs. Learning vocabulary and other language constructs as a use-case explored in this study, presents just one of many possibilities that offer venues for future work in AR classrooms as it is likely that the technology will get better and widely accessible in the future [58].

Acknowledgments

The authors wish to thank Cuauthli Campos for helping with preparing the video and all the volunteers who participated in the user study.

This research was supported by European Commission through the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund). We also acknowledge support from the Slovenian research agency ARRS (program no. BI-DE/20-21-002, P1-0383, J1-9186, J1-1715, J5-1796, and J1-1692).

References

[1] A taxonomy for learning, teaching, and assessing; a revision of bloom's taxonomy of educational objectives, complete edition, 2001.
[2] *Implicit emotional tagging of multimedia using EEG signals and brain computer interface*, 2009.
[3] *An Adaptive AR Tutor For Cabling a Network Topology.*, 2017.
[4] O. Alrashidi, H. P. Phan, and B. H. Ngu. Academic engagement: An overview of its definitions, dimensions, and major conceptualisations. *International Education Studies*, 9(12):41–52, 2016.
[5] American Institute of Physics. *What levels of guidance promote engaged exploration with interactive simulations?*, vol. 1064, 2008.

[6] M. Amiryousefi and S. Ketabi. Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching & Research*, 2(1), 2011.
[7] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent tutoring systems. *Science*, 228(4698):456–462, 1985.
[8] J. J. Appleton, S. L. Christenson, and M. J. Furlong. Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5):369–386, 2008.
[9] P. Arvanitis and P. Krystalli. Digital storytelling and augmented reality-based scenarios for foreign language teaching. In *Digital Pedagogies and the Transformation of Language Education*, pp. 1–22. IGI Global, 2021.
[10] D. Ary, L. C. Jacobs, C. K. S. Irvine, and D. Walker. *Introduction to research in education*. Cengage Learning, 2018.
[11] R. C. Atkinson. Mnemotechnics in second-language learning. *American psychologist*, 30(8):821, 1975.
[12] J. Birt, Z. Stromberga, M. Cowling, and C. Moro. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information*, 9(2):31, 2018.
[13] D. Boud, R. Cohen, and D. Walker. *Using experience for learning*. McGraw-Hill Education (UK), 1993.
[14] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
[15] J. S. Bruner. The act of discovery. *Harvard educational review*, 1961.
[16] A. I. Canhoto and J. Murphy. Learning from simulation design to develop better experiential learning initiatives: An integrative approach. *Journal of Marketing Education*, 38(2):98–106, 2016.
[17] K. J. Carbonneau and S. C. Marley. Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *The Journal of Experimental Education*, 83(4):495–513, 2015.
[18] R. C. Clark, F. Nguyen, and J. Sweller. *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons, 2011.
[19] R. E. Clark. How much and what type of guidance is optimal for learning from instruction. *Constructivist instruction: Success or failure*, pp. 158–183, 2009.
[20] A. D. Cohen and E. Aphek. Retention of second-language vocabulary overtime: Investigating the role of mnemonic associations. *System*, 8(3):221–235, 1980.
[21] J. Cohen. Statistical power analysis for the behavioral sciences. *England: Routledge*, 1988.
[22] L. J. Cronbach and R. E. Snow. *Aptitudes and instructional methods: A handbook for research on interactions.* Irvington, 1977.
[23] F.-A. Dita. A foreign language learning application using mobile augmented reality. *Informatica Economica*, 20(4):76, 2016.
[24] F. Draxler, A. Labrie, A. Schmidt, and L. L. Chuang. Augmented reality to enable users in learning case grammar from their real-world interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.
[25] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.
[26] N. C. Ellis. The psychology of foreign language vocabulary acquisition: Implications for call. *Computer Assisted Language Learning*, 8(2-3):103–128, 1995.
[27] Erasmus+. Factsheets and statistics on Erasmus+. `https://erasmus-plus.ec.europa.eu/node/2585`. 2022-05-24.
[28] G. Ferrero, F. Bichai, and M. Rusca. Experiential learning through role-playing: Enhancing stakeholder collaboration in water safety plans. *Water*, 10(2):227, 2018.
[29] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis. Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE transactions on Information Technology in Biomedicine*, 14(3):589–597, 2010.
[30] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109, 2004.
[31] P. A. Freund, J.-T. Kuhn, and H. Holling. Measuring current achievement motivation with the qcm: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, 51(5):629–634, 2011.
[32] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British journal of surgery*, 91(2):146–150, 2004.
[33] J.-Q. Guan, L.-H. Wang, Q. Chen, K. Jin, and G.-J. Hwang. Effects of a virtual reality-based pottery making approach on junior high school students' creativity and learning engagement. *Interactive Learning Environments*, pp. 1–17, 2021.
[34] A. K. Halabi. Applying an instructional learning efficiency model to determine the most efficient feedback for teaching introductory accounting. *Global Perspectives on Accounting Education*, 3(1):6, 2006.
[35] P. Herman and L. M. Gomez. Taking guided learning theory to school: Reconciling the cognitive, motivational, and social contexts of instruction. pp. 74–93, 2009.
[36] G. Huang, X. Qian, T. Wang, F. Patel, M. Sreeram, Y. Cao, K. Ramani, and A. J. Quinn. Adaptutar: An adaptive tutoring system for machine tasks in augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021.
[37] T.-C. Huang, C.-C. Chen, and Y.-W. Chou. Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, 96:72–82, 2016.
[38] A. Ibrahim, B. Huynh, J. Downey, T. Höllerer, D. Chun, and J. O'donovan. Arbis pictus: A study of vocabulary learning with augmented reality. *IEEE transactions on visualization and computer graphics*, 24(11):2867–2874, 2018.
[39] IEEE. *Remarks on emotion recognition from multi-modal bio-potential signals*, vol. 2, 2003.
[40] M. Jantjies, T. Moodley, and R. Maart. Experiential learning through virtual and augmented reality in higher education. In *Proceedings of the 2018 international conference on education technology management*, pp. 42–45, 2018.
[41] L. Jarmon, T. Traphagan, M. Mayrath, and A. Trivedi. Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in second life. *Computers & Education*, 53(1):169–182, 2009.
[42] L. Joplin. On defining experiential education. *Journal of experiential education*, 4(1):17–20, 1981.
[43] S. Joseph. *Measuring cognitive load: A comparison of self-report and physiological methods*. PhD thesis, Arizona State University, 2013.
[44] S. Kalyuga, P. Chandler, and J. Sweller. Learner experience and efficiency of instructional guidance. *Educational Psychology*, 21(1):5–23, 2001.
[45] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3):502–512, 2008.
[46] G. Kearsley and B. Shneiderman. Engagement theory: A framework for technology-based teaching and learning. *Educational technology*, 38(5):20–23, 1998.
[47] P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.
[48] K. R. Koedinger and V. Aleven. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3):239–264, 2007.
[49] R. A. L. Koelstra. Affective and implicit tagging using facial expressions and electroencephalography., 2012.
[50] D. Kolb. *Experiential learning: experience as the source of learning and development*. Prentice Hall, 1984.
[51] D. A. Kolb. Experiential learning theory and the learning style inventory: A reply to freedman and stumpf. *Academy of Management Review*, 6(2):289–296, 1981.
[52] D. A. Kolb, R. E. Boyatzis, C. Mainemelis, et al. Experiential learning theory: Previous research and new directions. *Perspectives on thinking, learning, and cognitive styles*, 1(8):227–247, 2001.
[53] G. F. Kuder and M. W. Richardson. The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160, 1937.
[54] D. Kuhn. Is direct instruction an answer to the right question? *Educational psychologist*, 42(2):109–113, 2007.
[55] A. W. Lazonder and R. Harmsen. Meta-analysis of inquiry-based learning: Effects of guidance. *Review of educational research*, 86(3):681–718, 2016.
[56] J. R. Lewis and J. Sauro. The factor structure of the system usability scale. In *International conference on human centered design*, pp. 94–103. Springer, 2009.

[57] L. H. Lewis and C. J. Williams. Experiential learning: Past and present. *New directions for adult and continuing education*, 1994(62):5–16, 1994.
[58] R. A. Liono, N. Amanda, A. Pratiwi, and A. A. Gunawan. A systematic literature review: learning with visual by the help of augmented reality helps students learn better. *Procedia Computer Science*, 179:144–152, 2021.
[59] P.-H. E. Liu and M.-K. Tsai. Using augmented-reality-based mobile learning material in efl english composition: An exploratory case study. *British journal of educational technology*, 44(1):E1–E4, 2013.
[60] S.-J. Lu and Y.-C. Liu. Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research*, 21(4):525–541, 2015.
[61] P. Mair and R. Wilcox. Robust statistical methods in r using the wrs2 package. *Behavior research methods*, 52(2):464–488, 2020.
[62] M. A. Mastropieri and T. E. Scruggs. *Teaching students ways to remember: Strategies for learning mnemonically.* Brookline Books, 1991.
[63] C. Mather, T. Barnett, V. Broucek, A. Saunders, D. Grattidge, and W. Huang. Helping hands: using augmented reality to provide remote guidance to health professionals. In *Context Sensitive Health Informatics: Redesigning Healthcare Work*, pp. 57–62. IOS Press, 2017.
[64] R. E. Mayer. Constructivism as a theory of learning versus constructivism as a prescription for instruction. *Constructivist instruction: Success or failure*, pp. 184–200, 2009.
[65] B. Mergel. Instructional design and learning theory, 1998.
[66] M. D. Merrill. First principles of instruction. *Educational technology research and development*, 50(3):43–59, 2002.
[67] N. Moorhouse, T. Jung, et al. Augmented reality to enhance the learning experience in cultural heritage tourism: An experiential learning cycle perspective. *eReview of Tourism Research*, 8, 2017.
[68] N. Moorhouse, M. C. tom Dieck, and T. Jung. An experiential view to children learning in museums with augmented reality. *Museum Management and Curatorship*, 34(4):402–418, 2019.
[69] A. Morrison, A. Oulasvirta, P. Peltonen, S. Lemmela, G. Jacucci, G. Reitmayr, J. Näsänen, and A. Juustila. Like bees around the hive: A comparative study of a mobile augmented reality map. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1889–1898. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518991
[70] D. Mueller and J. M. Ferreira. Marvel: A mixed-reality learning environment for vocational training in mechatronics. In *Proceedings of the Technology Enhanced Learning International Conference (TEL 03)*, 2003.
[71] F. G. Paas and J. J. Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743, 1993.
[72] A. Paivio and A. Desrochers. Mnemonic techniques in second-language learning. *Journal of Educational Psychology*, 73(6):780, 1981.
[73] R. L. Plackett. Karl pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique*, pp. 59–72, 1983.
[74] M. Pressley, J. R. Levin, and H. D. Delaney. The mnemonic keyword method. *Review of Educational Research*, 52(1):61–91, 1982.
[75] A. L. Putnam. Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, 1(2):130, 2015.
[76] M. R. Raugh and R. C. Atkinson. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 67(1):1, 1975.
[77] J. Reeve and C.-M. Tseng. Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4):257–267, 2011.
[78] F. Rheinberg, R. Vollmeyer, and B. D. Burns. Fam: Ein fragebogen zur erfassung aktueller motivation in lern-und leistungssituationen (langversion, 2001). *Diagnostica*, 2:57–66, 2001.
[79] J. T. Richardson and A. Jones. *Mental imagery and human memory*. Macmillan International Higher Education, 1980.
[80] R. Rzayev, S. Korbely, M. Maul, A. Schark, V. Schwind, and N. Henze. Effects of position and alignment of notifications on ar glasses during social interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. Association for Computing Machinery, New York, NY, USA, 2020.
[81] SAGE Publications Sage CA: Los Angeles, CA. *Engagement and Competence in VR and non-VR Environments*, vol. 61, 2017.
[82] G. Salomon. The differential investment of mental effort in learning from different sources. *Educational psychologist*, 18(1):42–50, 1983.
[83] W. B. Schaufeli. What is engagement. *Employee engagement in theory and practice*, 15:321, 2013.
[84] M. Schrepp. Ueq-user experience questionnaire, 2019.
[85] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *IJIMAI*, 4(6):103–108, 2017.
[86] H. Schuman, S. Presser, and J. Ludwig. Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, 45(2):216–223, 1981.
[87] P. Seedhouse, A. Preston, P. Olivier, J. Dan, H. Philip, B. Madeline, A. Rafiev, and M. Kipling. The european digital kitchen project. *Bellaterra Journal of Teaching & Learning Language & Literature*, 7(1):1–16, 2014.
[88] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.
[89] Springer. *Time-frequency analysis of EEG signals for human emotion detection*, 2008.
[90] L. P. Steffe and J. E. Gale. *Constructivism in education*. Psychology Press, 1995.
[91] G. M. Sullivan and A. R. Artino Jr. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542, 2013.
[92] L. Sullivan. Interquartile range (iqr), 2016.
[93] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 161–174, 2019.
[94] S. Tobias. When do instructional methods. *Educational Researcher*, 11(4):4–9, 1982.
[95] S. Tobias and T. M. Duffy. *Constructivist instruction: Success or failure?* Routledge, 2009.
[96] F. B. Topu and Y. Goktas. The effects of guided-unguided learning in 3d virtual environment on students' engagement and achievement. *Computers in Human Behavior*, 92:1–10, 2019.
[97] K. L. Vaughan, R. E. Vaughan, and J. M. Seeley. Experiential learning in soil science: Use of an augmented reality sandbox. *Natural Sciences Education*, 46(1):1–5, 2017.
[98] C. D. Vazquez, A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2172–2179, 2017.
[99] M. Weber, J. Giacomin, A. Malizia, L. Skrypchuk, V. Gkatzidou, and A. Mouzakitis. Investigation of the dependency of the drivers' emotional experience on different road types and driving conditions. *Transportation research part F: traffic psychology and behaviour*, 65:107–120, 2019.
[100] M. Weerasinghe, A. Quigley, J. Ducasse, K. Č. Pucihar, and M. Kljun. Educational augmented reality games. pp. 3–32, 2019.
[101] X. Wei, D. Weng, Y. Liu, and Y. Wang. A tour guiding system of historical relics based on augmented reality. In *2016 IEEE Virtual Reality (VR)*, pp. 307–308. IEEE, 2016.
[102] G. Westerfield, A. Mitrovic, and M. Billinghurst. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, 25(1):157–172, 2015.
[103] C. Weston and P. A. Cranton. Selecting instructional strategies. *The Journal of Higher Education*, 57(3):259–288, 1986.
[104] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015.
[105] A. F. Wise and K. O'Neill. Beyond more versus less: A reframing of the debate on instructional guidance. 2009.
[106] J. B. Worthen and R. R. Hunt. *Mnemonology: Mnemonics for the 21st century*. Psychology Press, 2011.
[107] S. Yang and B. Mei. Understanding learners' use of augmented reality in language learning: insights from a case study. *Journal of Education for Teaching*, 44(4):511–513, 2018.

Chapter 6

Publication 03: VocabulARy [2]

Title: VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations

Authors: Maheshya Weerasinghe, Verena Biener, Jens Grubert, Aaron Quigley, Alice Toniolo, Klen Čopič Pucihar, Matjaž Kljun

Year: 2022

Journal: IEEE Transactions on Visualization and Computer Graphics (Q1)

DOI: 10.1109/TVCG.2022.3203116

Link: `https://ieeexplore.ieee.org/document/9872027`

Summary: The VocabulARy Study, focused on investigating the impact of different *Display Configurations* (i.e., AR HMD and non-AR 2D display) and different types of *Guidance* (i.e., keyword only and keyword + visualisations) on user engagement and learning performance in a vocabulary learning task. We designed an AR vocabulary system using an HMD that visually annotated objects in the user's surroundings and compared it to a non-AR system on a tablet. This study directly addressed RQ1 (How can XR technology be used to create immersive and interactive experiential learning environments by exploring the importance and effectiveness of different configurations of XR, ExL and instructional guidance?), as well as RQ4 (How do the different configurations of XR and/or degree of dynamic adaptations of guidance affect the user experience, engagement and, consequently, the learning outcomes of immersive ExL environments?). This provided valuable insights into the importance and effectiveness of AR and different types of guidance in immersive vocabulary learning environments.

VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations

Maheshya Weerasinghe, *Student Member, IEEE*, Verena Biener, Jens Grubert, Aaron Quigley, Alice Toniolo, Klen Čopič Pucihar and Matjaž Kljun

Fig. 1. VocabulARy prototype. (a) Participant interacting with VocabulARy during the study; (b) VocabulARy prototype though HoloLens2 in KEYWORD + VISUALISATION instruction mode (the Japanese word "hagaki" sounds as the English phrase "hug a key" (keyword) and visualised with an animated hand grabbing a key (visualisation)); (c) Participant interacting with non-AR version of VocabulARy in KEYWORD instruction mode (Note there is no visualisation of the keyword). AR and non-AR condition were tested with both instruction modes.

Abstract—Learning vocabulary in a primary or secondary language is enhanced when we encounter words in context. This context can be afforded by the place or activity we are engaged with. Existing learning environments include formal learning, mnemonics, flashcards, use of a dictionary or thesaurus, all leading to practice with new words in context. In this work, we propose an enhancement to the language learning process by providing the user with words and learning tools in context, with VocabulARy. VocabulARy visually annotates objects in AR, in the user's surroundings, with the corresponding English (first language) and Japanese (second language) words to enhance the language learning process. In addition to the written and audio description of each word, we also present the user with a keyword and its visualisation to enhance memory retention. We evaluate our prototype by comparing it to an alternate AR system that does not show an additional visualisation of the keyword, and, also, we compare it to two non-AR systems on a tablet, one with and one without visualising the keyword. Our results indicate that AR outperforms the tablet system regarding immediate recall, mental effort and task-completion time. Additionally, the visualisation approach scored significantly higher than showing only the written keyword with respect to immediate and delayed recall and learning efficiency, mental effort and task-completion time.

Index Terms—Augmented Reality, vocabulary learning, keyword method, contextual learning

1 Introduction

Learning a language is a complex task that requires dedication, perseverance and hard work. The basic learning process consists of comprehension of input (i.e. hearing or reading), comprehensible output (i.e. speaking or writing) and feedback (i.e. identifying errors and making changes in response) [6, 45]. Through these processes we learn vocabulary and grammar enhancing our language comprehension and expression abilities.

Expanding one's vocabulary is an essential element of language learning and in vocabulary learning, methods for improving learners' memory play a vital role. Mnemonics is one such effective method, in which the learner attempts to link new learning with prior knowledge through the use of visual and/or acoustic cues. Keywords are one such practical technique in which the learner attempts to create a symbolic link between new and prior knowledge using associations triggered by keywords, a method shown to be particularly effective in prior research [4].

Furthermore, previous research shows that learning vocabulary can be enhanced through an encounter with words in context [50]. Existing learning environments include formal learning, flashcards, use of a dictionary or thesaurus, all leading to practice with new words in context. For example, in formal learning the context is provided by the instructor or the provided instructional materials, in flashcards it is formed through images depicted on physical cards, in thesaurus it is provided through the provision of synonyms.

Consumer devices such as smartphones, tablets and head mounted displays can be used to enhance existing learning environments or to provide new ones. These systems enable technology driven paradigm shifts such as e-learning [40, 41, 44], and more recently m-learning (mobile learning) [19, 32, 71]. All are capable of enhancement through better provision of learning context and methods for improving learners' memory. Furthermore, these systems are also capable of running Augmented Reality (AR) applications which have the potential to make language learning more intuitive and immersive because of their intrinsic ability to visualise digital information within a real world context.

- *Maheshya Weerasinghe is with the University of Primorska, Slovenia and the University of St Andrews, United Kingdom. E-mail: amw31@st-andrews.ac.uk.*
- *Verena Biener and Jens Grubert are with the Coburg University of Applied Sciences, Germany. E-mail: {jens.grubert | verena.biener}@hs-coburg.de.*
- *Aaron Quigley is with the University of New South Wales, Australia. E-mail: a.quigley@unsw.edu.au.*
- *Alice Toniolo is with the University of St Andrews, United Kingdom. E-mail: a.toniolo@st-andrews.ac.uk.*
- *Klen Čopič Pucihar and Matjaž Kljun are with the University of Primorska, Slovenia. E-mail: {klen.copic | matjaz.kljun}@famnit.upr.si.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

This is particularly important for vocabulary learning because it allows word encounters in real-world context, an important catalyst for vocabulary learning [28, 62, 76].

Despite the fact that prior work looked at AR for vocabulary learning discovering several benefits, such as better improved retention, higher enjoyment, motivation and engagement, none provide a direct comparison of AR applications that run in head-mounted displays to the same technique within a non-AR interface. Furthermore, to the best of our knowledge, no existing evaluation of vocabulary learning that combines keywords with visualisations exist.

This paper contributes to addressing this gap with VocabulARy, an AR application for vocabulary learning that visually annotates objects in AR, in the user's surroundings, with the corresponding English (first language) and Japanese (second language) words. In addition to the written and audio description of each word, VocabulARy also presents the user with a keyword and its visualisation to enhance memory retention. We evaluate the VocabulARy prototype by comparing it to an alternate AR system that does not show an additional visualisation of the keyword and also, we compare it to two non-AR systems on a tablet, one with and one without visualising the keyword. The results show that AR outperforms the NON-AR (tablet) system regarding short-term retention, mental effort and task-completion time. Additionally, the visualisation approach scored significantly higher than only showing the written keyword with respect to immediate and delayed recall and learning efficiency, mental effort and task-completion time.

2 RELATED WORK

Vocabulary learning can be enhanced through methods for improving learners' memory [39, 53] or through an encounter with words in context [50]. AR is an emerging technology for learning in real-world context and to scaffold this we structure our related work into: Learning context, Vocabulary learning in AR and Memory enhancement techniques. To better position our work in the context of language learning in AR environments we also classify prior work based on AR devices, learning content, presentation and learning method (Table 1).

Table 1. Selected prior work related to language learning in AR environments.

Study	AR Device	Content	Presentation	Learning Method
Draxler et al. (2020)	Hand-held	Grammar	Visual, Audio & Text	Context-based
Dalim et al. (2020)	Desktop	Vocabualry	Visual, Audio & Text	Experiential
Arvanitis et al. (2020)	Hand-held	Vocabualry	Visual, Audio & Text	Self-directed
Ibrahim et al. (2018)	HMD	Vocabualry	Visual, Audio & Text	Context-based
Yang & Mei (2018)	Hand-held	Vocabualry	Visual, Audio & Text	Context-based
Hautasaari et al. (2019)	Hand-held	Vocabualry	Audio	Context-based
Vazquez et al. (2017)	HMD	Vocabualry	Visual, Audio & Text	Context-based
Santos et al. (2016)	Hand-held	Vocabualry	Visual, Audio & Text	Context-based
Dita (2016)	Hand-held	Vocabualry	Visual & Text	Context-based
Li et al. (2014)	Hand-held	Vocabualry	Visual & Text	Not Speicified
Liu & Tsai (2013)	Hand-held	Vocabualry	Visual & Text	Context-based
VocabulARy	HMD	Vocabualry	Visual, Audio & Text	Context-based & **Keyword**

2.1 Learning Context

It has been shown that people are more motivated to learn, if they can see the importance of the content with respect to the situation or, if they find the content interesting [49]. For example, being in a bar in a foreign country is likely to increase the interest in learning words and sentences required for ordering a coffee. Additionally, the context makes it possible to form associations that help later retrieval in similar circumstances [28, 62, 76, 80]. In other words, new words relevant to the learning context are more likely to be recalled than unrelated words [15].

AR has the ability to provide context-specific information in an interactive manner. In addition, AR can take any situation, location, environment, or experience to a whole new level by combining digital information with real-world contents. Thus, it has the potential to create more engaging and immersive learning environments. There exists a considerable body of previous work on AR systems that support learning in real-world contexts. For instance, there are systems that provide labels of new words corresponding to real-world objects [13, 62], while others create imaginary settings to describe and enhance the physical properties of everyday objects [27, 73].

2.2 Vocabulary Learning in AR

Previous studies have shown that AR offers many advantages for language and vocabulary learning. For instance, some studies reported that AR improved learning achievements and boosted motivation, engagement and collaboration among learners [12, 13, 25, 28, 62]. Despite, some technical limitations of using AR for learning should be taken into account such as such as educators' limited proficiency with the relatively new technology [29] or the trade-off between connecting the experience to the context of the current location and providing a flexible and portable experience [79].

Fujimoto et al. [18] have shown that users can memorise AR information better if it is shown within the location of a target object in the real world (e.g. AR information about a country shown over a map within this country). However, the information to be memorised in study did not take the context of the real environment and users' surroundings into account.

Several studies presented applications for learning vocabulary using hand-held AR devices. Hautasaari et al. [25] developed the VocaBura smartphone application for learning vocabulary during dead time. The application tracks a users' GPS locations and presents vocabulary related to the current location via audio. A study comparing this to an audio-only method showed that 7 days after the study, participants could recall significantly more words. Santos et al. [62] presented a handheld AR system that displays text, images, animation and sound next to corresponding real-world objects to learn Filipino or German. They compared this system to a non-AR tablet application using a flash card method. Their results indicate that for tests directly after the experiment non-AR users performed better, yet this difference was not detected for long-term retention.

Positive effects of AR technology in the context of increased motivation and enjoyment have also been detected. For example, Dalim et al. [12] presented a system combining desktop-AR and speech recognition (TeachAR) and found that it increases children's knowledge gain and enjoyment. Similarly, Li et al. [34] explored an AR application for language learning and found that it increased motivation in the beginning, yet for most participants motivation decreased at the end of the study.

The existing body of literature also includes applications for learning vocabulary on AR head-mounted devices. While most previous systems used some sort of marker to align virtual content with physical objects, Vazquez et al. [76] presented a platform (WordSense) that detects objects in the physical environment and augments them with additional content for language learning including words, sentences, definitions, videos and audio. However, no formal user study was conducted to evaluate the system.

Another example of using AR head-mounted devices for language learning is ARbis Pictus, a system presented by Ibrahim et al. [28] which labels objects in the user environment with the corresponding vocabulary in the target language. They compared this system to a conventional flashcard-based system and found that AR was more effective and enjoyable and participants could remember words better both shortly after the experiment and four days later. However, the significance of these findings is limited, because the flashcard and AR systems were inherently different. For example, with flashcards the word was shown on the opposite side to the image depicting word meaning thus the image and word were never shown together. This was not the case in the AR condition where word annotations were always visible for all objects in the scene. Therefore it is not clear if AR accounted for the improved performance or the different learning approach.

In contrast to the presented studies we compare our AR prototype to a non-AR system that is as similar as possible to enable us to measure only the effect of AR without confounding variables like the learning method. To our knowledge, such an experiment has not yet been conducted for AR applications that run on head-mounted devices.

2.3 Memory Enhancement Techniques

Research on memory and learning has shown that learning performance and retention depend on different strategies and techniques that can be used to process information in learning [14]. "Mnemonic" is one such technique where the memory capabilities are enhanced by connecting new information to prior knowledge through the use of visuals and/or acoustic cues [39, 53]. Several researchers experimentally showed that "Mnemonic" techniques improve memory and recall, especially in the area of language learning [1, 10, 47, 54].

In the field of language learning, mnemonics have mostly been used for vocabulary learning [2]. One such mnemonic method is the "keyword method" in which learners connect the sound of a word they want to learn to one they already know in either their first language or the target language. Through this process learners create a mental image that helps them remember the association [50]. For example, the Japanese word for bread is "sokupan" which in English sounds like "Sock + Pan". As a result, the learner can imagine a sentence that links a mnemonic keyword with the foreign word. For example, "sokupan" can be imagined as frying a sock and putting it on a slice of bread.

A wide range of existing studies in the broader literature have explored the effectiveness of the keyword method [2, 4, 54, 77]. In this context, comparing the keyword method against other methods in vocabulary learning is one of the most common research designs. There, the keyword method has been compared with learning words in context or learning words with no given strategy. For example, Pressley et al. [51] found the keyword method to be significantly more effective in learning over the context method. Also, Sagarra and Alba [58] compared rehearsal, semantic mapping displays and the keyword method, and found that the keyword method resulted in the best retention. It has also been shown that the keyword method is superior over systematic teaching [30, 51]. In 1975, Atkinson and Raugh [4] found that participants who were given a keyword along with the translation learned more words and also remembered more words after 6 weeks. In the same sense, Raug et al. [55] evaluated the use of the keyword method over a long period of 8 to 10 weeks to teach Russian vocabulary and found it to be highly effective.

Altogether, a significant number of research studies have shown that the keyword method of vocabulary learning is highly effective, yet others showed mixed results [48, 77]. For example, a study conducted by Zheng Wei [77] found no significant differences between the keyword method, the word-part technique (recognizing part of a word) and the self-strategy. From the perspective of the learning method, VocabulARy builds upon the work of Anonthanasap et al. [2] in which the authors propose an interactive vocabulary learning system to teach Japanese that automatically creates keywords using phonetic algorithms. There, if the learner selects an image in the system, the phonetically similar words with image representations will gather around the selected image. Results showed that the keyword-based vocabulary learning system required significantly lower workload than the other compared methods (e.g. paper dictionary and static visualisation in a form of an image).

In summary, our work was inspired by Santos et al. [62], Vazquez et al. [76] and Ibrahim et al. [28] who already used AR devices to augment real world objects with annotations for vocabulary learning. We combined this approach of providing context with a keyword method which has proven to work well in various experiments thus far [2, 4, 58]. To further advance this learning method we augment keywords with visualisations. AR provides ideal conditions for that, because the keyword and its visualisation can be shown in the same context as the corresponding physical object. However, in contrast to existing visualisations of keyword approaches we make a careful selection of keywords so that they not only sound similar, but can also be visualised with an animation in a meaningful way. For example, a Japanese word for bread is "sokupan" and sounds similar to keyword "Sock+Pan" which can be visualised with an animation of frying a sock in a pan and putting it on a slice of bread. We do this to uncover if one can improve vocabulary learning beyond the influence of the traditional keyword method, by augmenting the keywords with animated visualisations. According to Shapiro and Waters [67] the level of visual imagery of a word enhances vocabulary learning suggesting our approach should work, however no formal evaluation of this exists (Table 1). This makes our work both ground breaking and timely.

3 Augmented Vocabulary Learning / Learning Vocabulary with Visualisations in Real Life Context

This work presents two prototype systems for vocabulary learning developed on an AR head-mounted-display (HMD) (i.e. Microsoft HoloLens 2) and an 10.5 *in* Android tablet device (i.e. Samsung Galaxy Tab S4) (Fig. 1). To the best of our knowledge, AR HMD systems for vocabulary learning have not yet been evaluated against comparable non-AR systems (see Sect. 2.2). Both AR and tablet systems combine the keyword method with physical objects. With the AR HMD, our system allows the user to look around a physical room where certain objects are labelled with a button indicating that their translation is available. Upon clicking these buttons with a hand gesture, the English and Japanese word, as well as a keyword with or without visualisation, appears. In addition to the words in both languages and a keyword, an audio of the pronunciation is played. The details of application design and implementation of both prototypes are described in the following sub sections.

3.1 Application Design

In this section we describe key design decisions regarding annotations, animated visualisations and interactions. Careful consideration was given to the selection of annotation and visualisation size. Previous research showed the size of images affects our ability to remember image content during naturalistic exploration [38]. However in such exploration individuals are first asked to freely explore an image without any instructions and are then asked about the details observed. As such there is no guarantee that the visual attention is equally distributed across the observed image and as the image gets smaller so does the key information, which makes it easier to miss it.

To prevent participants from missing key information we design our application to effectively guide visual attention. The application shows only one word visualisation at a time, which avoids cluttering the scene and overloading participants with too much information. Furthermore, we specifically chose to place AR buttons on the physical surface at close proximity to the object of which word was being memorised. This led users to the appropriate physical location from which AR visualisations are clearly visible as the corresponding annotation and animated visualisation size were appropriated for such viewing. To make the NON-AR and AR condition as comparable as possible we made sure that the relative size of annotations and animation was roughly the same in both conditions.

Furthermore, in one of the instruction modes, the keyword is also accompanied by its animated visualisation. Such a visualisation consists of a 3D model resembling the keyword and a short animation involving the objects in question. Besides, the user can also listen to the pronunciation of the Japanese word again by clicking a virtual button that is displayed next to the keyword. For example, Figure 1b shows how the English word "Postcard" and the corresponding Japanese word "Hagaki" are displayed. "Hagaki" sounds like "Hug + A Key", so it is displayed as a keyword and visualised through "hugging a key".

3.2 Implementation of AR Prototype

The AR prototype was implemented for Microsoft HoloLens 2 [42] using the Unity3D game development engine [75]. For camera pose tracking, the HoloLens inbuilt tracking system was used. To initialise the positions of augmentations in our application, we used Vuforia [52] and our custom-made image markers (see Figure 2). We opted for markers in order to support a reliable and accurate detection of physical objects that correspond to the set of vocabulary. These markers were removed from the scene after initialisation.

It would be technically possible to use object recognition techniques to perform object identification and localisation as in [7, 59, 60]. Such implementation could support arbitrary environments without prior preparation, which would enable wide implementation of the system.

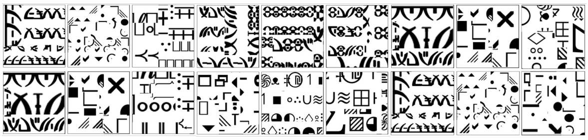


Fig. 2. Custom-made image markers.

However, this was not the scope of this work, as we focus on the effect of learning vocabulary using AR and visualisations.

To interact with the virtual contents, we use HoloLens' built-in hand-tracking and gesture inputs, which allow the user to interact with virtual content by moving the hands or fingers to content's corresponding positions. More specifically we chose to use virtual buttons placed on top of planar physical surfaces such as a table or a wall. In this way, touching a surface acted as a tangible feedback making the button press more realistic.

3.3 Implementation of Android Prototype

The Android version of the prototype was also implemented using the Unity 3D development environment, but deployed on a Samsung Galaxy Tab S4 [36] tablet device. Its functionality is similar to the AR prototype. However, instead of seeing the real world environment, an image of an environment is displayed on the screen. In our prototype this was either a kitchen or an office environment. As in the AR prototype certain objects are accompanied with a button. If the user touches the button, the corresponding English and Japanese words, keywords, and animated visualisations appear (Figure 1c). Visualisations only appear in one instruction mode. As the size of all objects was adapted to be clearly visible on the screen and to ensure that the relative size of annotations and animation was roughly the same in both conditions (see Sect. 3.1), we did not provide a feature to zoom into the scene.

Compared to the AR implementation, the tablet application can be used anywhere as it can also show scenes that are not related to the real-world environment around the learner. Because all kinds of virtual scenes can be presented, the tablet allows a more flexible use, such as learning words related to a forest while sitting in the living room.

3.4 Generating Keywords

For generating the keywords, we conducted a small informal survey with 7 participants. They were presented with 28 Japanese-English word pairs and were asked to come up with English words sounding similar to the Japanese words. At the end, we selected 20 words for the study for which the participants could come up with good keywords. As already mentioned, the process of finding keywords could also be automatized [2], but for the scope of this study, this was not needed.

4 RESEARCH METHOD

This section describes the study conditions, study design, participants' sampling, study procedure, data collection instruments, and analysis.

4.1 Study Conditions

We designed four study conditions based on two distinct vocabulary learning scenarios. The first scenario displays ten physical objects related to a kitchen environment, while the second shows an office environment with ten relevant physical objects. In each of these scenarios a different INSTRUCTION MODE is provided. This is either a KEYWORD instruction mode or a KEYWORD + VISUALISATION instruction mode. In the KEYWORD condition, only the written keywords are displayed to support the participants in remembering the word. In the KEYWORD + VISUALISATION condition, an animated 3D visualisation of the keyword is displayed in addition to the written keyword. These variations are presented to the participants on two different INTERFACES, one in AR on a HoloLens2 and one in NON-AR on an Android tablet. These study conditions are illustrated in Figure 3.

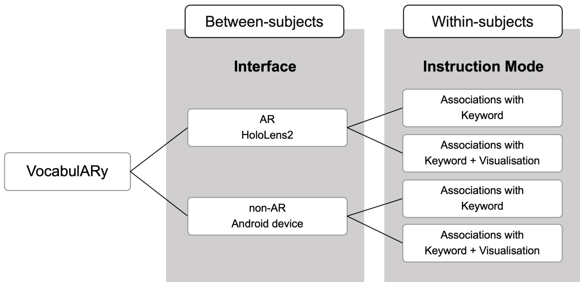


Fig. 3. Study design and conditions.

4.2 Study Design

Our study design has two independent variables: INSTRUCTION MODE which is KEYWORD or KEYWORD + VISUALISATION and INTERFACE which is either AR or NON-AR. We used a 2×2 mixed design (see Figure 3) because a within-subjects design would make the study, which is mentally demanding, too long (i.e. approximately two hours). We believe that such a long duration could intensify the fatigue and hinder the performance of the participants. Furthermore, running all 4 conditions in a within-subject design would require 2 additional learning scenarios making it more difficult to counterbalance for scenario effects. Reducing the study length by running the study in several sessions also introduces other biases and practical issues. Therefore, the INSTRUCTION MODE was evaluated as a within-subjects variable and the INTERFACE as a between-subjects variable. This means each participant was either using the AR-prototype or the NON-AR-prototype, but all participants experienced the KEYWORD and the KEYWORD + VISUALISATION conditions.

To avoid the "order effects", which may have an influence on participants' performance due to the order in which the conditions are presented [66], the order of the INSTRUCTION MODE as well as the order of the learning scenarios (the kitchen and the office environments) was counter balanced. Special care was given to counterbalance the learning scenario across all independent variables.

4.3 Participants

The study was completed by 32 participants, all voluntarily recruited from our university. None of the participants had any prior knowledge of the Japanese language (identified via a short competency test questionnaire). The between subject sample comprised of 16 participants for the AR condition (10 females) and 16 participants for the NON-AR condition (7 females). All the participants were between the age of 19 to 30 years, with the mean of $\bar{x} = 21.6$ and $SD = 2.1$.

All our participants were computer science undergraduate and graduate students. No student had previous experience with AR HMDs. The mean age for the AR group was $\bar{x} = 22.13$ ($SD = 2.68$), and for NON-AR group $\bar{x} = 21.13$ ($SD = 1.26$). The percentage of females in the AR group was $\bar{x} = 62.5\%$, and in the NON-AR group $\bar{x} = 43.75\%$. The groups were comparable.

4.4 Procedure

On arrival participants were first randomly assigned to one of the two groups (AR or NON-AR). Next, we randomly selected which instruction mode would be used first (KEYWORD or KEYWORD + VISUALISATION). Finally, the learning scenario was also randomly chosen (kitchen or office environment). All randomisations were counterbalanced.

After being assigned to a particular condition, participants were given a consent form to sign, together with the Participant Information Sheet (PIS) outlining the entire research process in simple language. After briefly explaining the vocabulary learning task with the two learning scenarios, they were asked to fill in the Questionnaire on Current Motivation (QCM) [57].

Before starting the actual task they completed a five-minute training session on a separate demo application to understand the VocabulARy interface. Participants were then given up to 15 minutes to complete the first language learning scenario with 10 words. After finishing the first scenario, they filled in the NASA Task Load Index (NASA-TLX) questionnaire [24] and, then answered a post-test questionnaire developed by the researchers to assess their immediate recall performance. After taking a 5 minutes break, they were again given up to 15 minutes to complete the second language learning scenario with 10 words. Subsequent to the second scenario, they filled in the same NASA-TLX and the immediate recall questionnaires.

After finishing the experiment, participants were given another two standard questionnaires – a system usability (SUS) [33] and a user experience questionnaire (UEQ) [64]. At the end, participants filled in a short post-questionnaire with demographic questions, questions about previous experience with AR technology and vision problems they might have. The entire experiment took 45 to 60 minutes.

One week later, participants were again requested to answer the same post-test questionnaire developed by the researchers to assess their delayed recall performance as undertaken in prior work [25].

4.5 Data Collection

In order to measure the task completion time, the time stamp data (start time and end time) were logged by the system. To measure the motivation, the short form of Questionnaire on Current Motivation (QCM) with 12 items/questions [17, 56] was used. Anxiety, challenge, interest, and probability of success were measured on a five-point Likert scale, with the labels "strongly disagree" at 1 and "strongly agree" at 5. Rather than aiming for constructing sub-dimensions (i.e., anxiety, challenge, interest, and probability of success), we used the mean score of the 12 items as an indicator for the overall motivation.

The NASA Task Load Index (NASATLX) [23, 43] was used to measure participants' subjective level of workload/mental effort. Participants rated five of its six dimensions (mental demand, physical demand, temporal demand, effort and frustration) on a 20-point scale ranging from 0 (very low) to 20 (very high). The endpoints of the sixth dimension (own performance) were success and failure. Finally, the overall workload/mental effort was calculated across these six dimensions.

In the retention questionnaires, participants were asked for the Japanese translations of the vocabulary they learned. This was undertaken immediately after interacting with the prototype (Immediate Recall) and after one week (Delayed Recall).

Learning efficiency was determined based on the ratio of performance to the difficulty of the learning task as proposed in [46]. The performance of each study condition was based on the recall scores participants obtained after completing the task. The difficulty of the task was based on the mental effort they invested during the learning phase (see Sect. 5.3). Performance and task difficulty data were then standardised using Formula 1 where z = Z-score, r = Raw data score, M = Population mean, and SD = Standard deviation.

$$z = \frac{r - M}{SD} \tag{1}$$

Next, the learning efficiency (E) was assessed for each of the four study conditions (Sect. 4.1) using Formula 2 [9, 22, 46] where E = Learning efficiency, z_P = Average performance in Z-scores, and z_M = Average task difficulty in Z-scores. This was done for both immediate recall performance (immediately after participants had completed the task) and delayed recall performance (a week after participants had completed the task). Note that square root of 2 in this formula comes from the general formula for the calculation of distance from a point, $p(x,y)$, to a line, $ax + by + c = 0$.

$$E = \frac{z_P - z_M}{\sqrt{2}} \tag{2}$$

To measure the usability of the system, we used the System Usability Scale (SUS), a ten question questionnaire originally created by Brooke, 1996 [5], on a five-point Likert scale, ranging from "Strongly" agree at 1 to "Strongly disagree" at 5. For measuring the user experience we used the short version of the User Experience Questionnaire (UEQ-S) [63, 64] with eight items/questions, reported on a 7-point Likert scale. The first four represent pragmatic qualities (Perspicuity, Efficiency and Dependability) and the last four hedonic qualities (Stimulation and Novelty) [63].

4.6 Data Analysis

The analysis was completed in R studio. Each data set collected in the study was first checked for mixed ANOVA assumptions. The normality assumption was checked using the Shapiro–Wilk normality test [68]. Most of the data sets were normally distributed with some of them only approximately normally distributed. The homogeneity of variance assumption of the between-subject factor (INTERFACE) was checked using the Levene's test [65] that confirmed homogeneity of variances for each variable ($p > 0.05$). Finally, the homogeneity of covariances of the between-subject factor (INTERFACE) was evaluated using the Box's M-test of equality of covariance matrices. The test showed homogeneity of covariances for each variable ($p > 0.001$). Considering the fact that some of the data sets were only approximately normally distributed, we used robust statistical methods implemented in the "WRS2" R package to conduct the analysis [37], which is a standard practice in such cases.

In all statistical analyses we used a significance level $p - value > 0.05$ and a restrictive confidence interval (CI) of 95%. For immediate recall, delayed recall, mental effort, task completion time and learning efficiency, the statistical significance was examined using a robust two-way mixed ANOVA on the 20% trimmed means–"bwtrim" [37].

For motivation, system usability and user experience, the statistical significance was examined using a Mann–Whitney U test [74]. Asterisk notation is used in tables to visualise statistical significance (ns: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$).

To assess the reliability of motivation and mental effort questionnaires, we performed a Cronbach's alpha test. Estimated reliability for each questionnaire (motivation Cronbach's $\alpha = 0.77$ and mental effort $\alpha = 0.85$) is acceptable for research purposes [3]. To measure the reliability of retention questionnaires, we conducted a Kuder-Richardson 20 test [31]. The $KR = 0.61 > 0.5$ value indicates that the reliability of the retention questionnaire is also acceptable.

We also conducted a power analysis to check and validate the results and findings of the study. We calculated the effect size (Cohen's d) for each data set collected [11], selected the minimum effect size (Cohen's $d = 0.69$) and estimated the statistical power $(1 - \beta)$ of data to check whether the type II error probability (β) is within an acceptable range for a given sample size (n = 16 per group) and a significance level ($\alpha = 0.05$). The estimated power value 0.96 shows that with the given sample size, we can have more than a 90% chance that we correctly reject the null hypothesis with a significance level of 0.05.

5 RESULTS

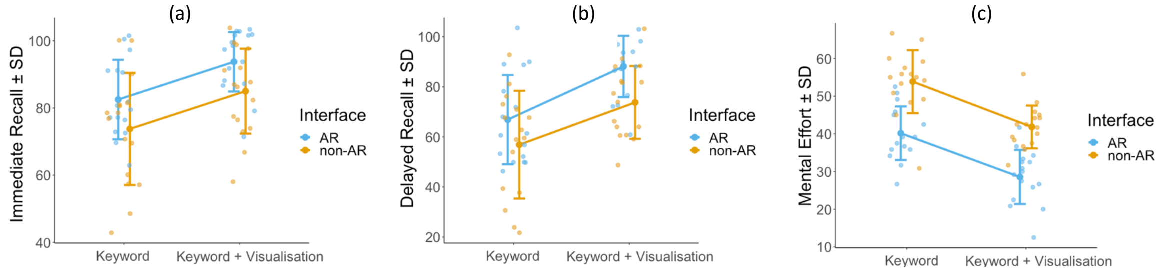
The results and analysis are based on the 32 participants who had completed all the facets of the study, i.e., the motivation questionnaire, the basic training, the vocabulary learning task and the post-questionnaires include mental effort, immediate recall and delayed recall tests. Participants had not undertaken any extra study for the delayed recall test.

We conducted a statistical analysis including gender as a between-subject factor (2 (GENDER) x 2 (INTERFACE) x 2 (INSTRUCTION MODE) mixed design) in order to exclude possible gender-based differences. The results did not indicate any statistical significant effect of GENDER on any dependant variable: immediate and delayed recall, mental effort, task completion time and, immediate and delayed learning efficiency. The results related to these variables are presented in the following subsections.

5.1 Immediate Recall

The mean values of immediate recall performance and the ANOVA results across study conditions, i.e., the INTERFACE (AR and NON-AR) and, the INSTRUCTION MODE (KEYWORD and KEYWORD+VISUALISATION) are shown in Figure 4a.

A significant main effect of INTERFACE on immediate recall performance could be detected ($F(1,60) = 7.46$, $p < 0.05$, $n^2 p = 0.11$). Here, participants' immediate recall scores were significantly better



Immediate Recall	SS	df	MS	F	p	n^2p
Interface	1225	1	1225	7.46	0.01 **	0.11
Instruction Mode	2025	1	2025	12.34	0.00 ***	0.17
Interface * Instruction Mode	0	1	0	0.00	1.00	0.00

Delayed Recall	SS	df	MS	F	p	n^2p
Interface	77	1	77	3.86	0.65	0.00
Instruction Mode	5814	1	5814	20.39	0.00 ***	0.25
Interface * Instruction Mode	76	1	76	0.27	0.61	0.00

Mental Effort	SS	df	MS	F	p	n^2p
Interface	2010	1	2910	57.05	0.00 ***	0.49
Instruction Mode	2236	1	2236	43.83	0.00 ***	0.42
Interface * Instruction Mode	0.72	1	0.72	0.01	0.906	0.00

Fig. 4. Means with standard deviation and ANOVA results for: (a) immediate recall performance in percentage of correctly remembered words; (b) delayed recall performance in percentage of correctly remembered words; (c) Mental effort.

in AR condition ($\bar{x} = 88.13\%$, $SD = 10.34$) compared to the NON-AR condition ($\bar{x} = 79.38\%$, $SD = 14.67$). Also, a significant main effect of INSTRUCTION MODE on immediate recall performance could be detected ($F(1,60) = 12.34$, $p < 0.001$, $n^2p = 0.17$). Results indicated that participants' immediate recall scores in KEYWORD + VISUALISATION ($\bar{x} = 89.38\%$, $SD = 10.75$) were significantly better than in KEYWORD ($\bar{x} = 78.13\%$, $SD = 14.26$). No significant interaction effect could be found between INTERFACE and INSTRUCTION MODE ($F(1,60) < 0.001$, $p > 0.05$, $n^2p < 0.001$).

5.2 Delayed Recall

The mean values of delayed recall performance and the ANOVA results across all study conditions, i.e., the INTERFACE (AR and NON-AR) and, the INSTRUCTION MODE (KEYWORD and KEYWORD + VISUALISATION) are shown in Figure 4b.

No significant main effect was found between INTERFACE on delayed recall performance ($F(1,60) = 3.86$, $p > 0.05$, $n^2p < 0.001$). The significance was only marginally missed. In addition, the mean values for the AR condition ($\bar{x} = 77.50\%$, $SD = 18.50$) were higher than for the NON-AR condition ($\bar{x} = 65.30\%$, $SD = 20.00$).

A significant main effect of INSTRUCTION MODE on delayed recall could be detected ($F(1,60) = 20.39$, $p < 0.001$, $n^2p = 0.25$). Results indicate that participants' delayed recall scores in KEYWORD + VISUALISATION ($\bar{x} = 80.88\%$, $SD = 13.60$) were significantly better than in KEYWORD ($\bar{x} = 61.88\%$, $SD = 19.65$). No significant interaction effect could be found between INTERFACE and INSTRUCTION MODE ($F(1,60) = 0.27$, $p > 0.05$, $n^2p < 0.001$).

5.3 Mental Effort

The mean values of mental effort (measured by NASA-TLX) invested to carry out the learning task and the ANOVA results over the study conditions are illustrated in Figure 4c.

A significant main effect of INTERFACE on mental effort could be detected ($F(1,60) = 57.05$, $p < 0.001$, $n^2p = 0.49$), such that the mental effort was significantly lower for AR condition ($\bar{x} = 34.36$, $SD = 7.14$) compared to the NON-AR condition ($\bar{x} = 47.85$, $SD = 7.02$). Also, a significant main effect of INSTRUCTION MODE on mental effort could be detected ($F(1,60) = 43.83$, $p < 0.001$, $n^2p = 0.42$). Here, participants' mental effort in KEYWORD + VISUALISATION ($\bar{x} = 35.19$, $SD = 6.42$) was significantly lower than in KEYWORD ($\bar{x} = 47.01$, $SD = 7.73$). No significant interaction effects was found between INTERFACE and INSTRUCTION MODE ($F(1,60) = 0.01$, $p > 0.05$, $n^2p < 0.001$).

5.4 Motivation

The mean values of motivation between INTERFACES (AR and NON-AR) is illustrated in Figure 5a. The data summarised in Figure 5a is analysed using a Mann–Whitney U test.

A significant effect was found between INTERFACES for participants' motivation ($U(N_{AR} = 16, N_{non-AR} = 16) = 48.50$, $p < 0.001$). There, the motivation was significantly higher for the AR condition ($\bar{x} = 3.69$, $SD = 0.36$) compared to the NON-AR condition ($\bar{x} = 3.29$, $SD = 0.27$).

5.5 System Usability

The answers to System Usability Scale (SUS) questions/items are reported on a 5-point Likert scale. The SUS scores are calculated as follows: for each of the odd numbered questions, subtract one from the user response, while for each of the even numbered questions, subtract their response from five and, add up the converted responses for each user and multiply that total by 2.5. This converts possible values to the range of 0 to 100 instead of 0 to 40. These adjustments are kept throughout the rest of the analysis.

The average SUS scores for AR and NON-AR INTERFACES are illustrated in Figure 5b. A Mann–Whitney U test indicated that there was no significant effect between the INTERFACES for SUS ($U(N_{AR} = 16, N_{non-AR} = 16) = 91.50$, $p > 0.05$).

5.6 User Experience

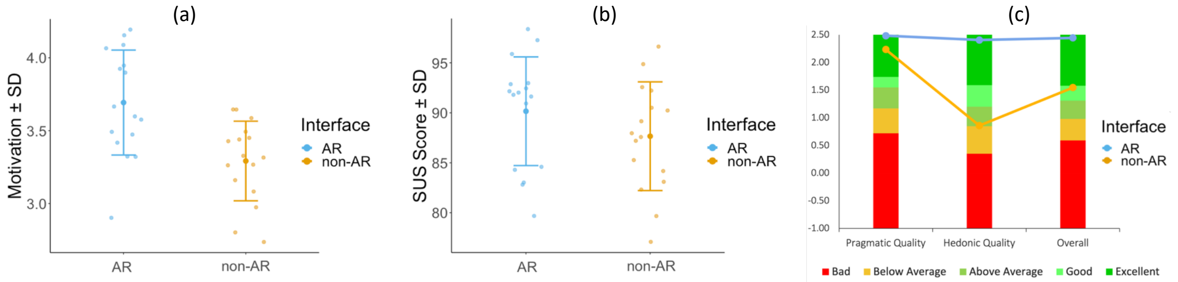
The UEQ-s questionnaire provides a benchmark to compare user experience between different systems [26]. It measures pragmatic qualities of a system (including efficiency, perspicuity and dependability) and hedonic qualities (including stimulation and novelty). The overall value was calculated from all 5 UEQ scale means. We adapted the standard method as suggested in [63,64] for calculating the scale means for each factor individually (efficiency, perspicuity, dependability, stimulation and novelty) and to obtain values for pragmatic quality, hedonic quality and overall user experience for both AR and NON-AR systems.

In the AR condition the pragmatic ($\bar{x} = 2.48$) and hedonic ($\bar{x} = 2.41$) qualities, as well as overall user experience ($\bar{x} = 2.45$) were all perceived as excellent (benchmarks for an excellent score: pragmatic > 1.73, hedonic > 1.55, overall > 1.58). In the NON-AR condition the mean value of pragmatic quality was perceived as excellent ($\bar{x} = 2.23$), and the overall experience as good ($\bar{x} = 1.55$) (benchmarks for a good score: pragmatic between 1.55 - 1.73, hedonic between 1.25 - 1.55, overall between 1.4 - 1.58). However, the hedonic factor was perceived as below average ($\bar{x} = 0.88$) (benchmarks for below average score: pragmatic between 0.73 - 1.14, hedonic between 0.88 - 1.24, overall between 0.68 - 1.01).

The overall user experience was analysed using a Mann–Whitney U test. The data is presented in Fig. 5c. A significant effect was found between INTERFACES ($U(N_{AR} = 16, N_{non-AR} = 16) = 16.00$, $p < 0.001$).

5.7 Task Completion Time

The mean values of task completion time across study conditions, i.e., the INTERFACE (AR and NON-AR) and, the INSTRUCTION MODE



Motivation			
	Statistics	p	Sample Size
Interface	48.50	0.00 ***	16

System Usability			
	Statistics	p	Sample Size
Interface	91.50	0.17	16

User Experience			
	Statistics	p	Sample Size
Interface	16.00	0.00 ***	16

Fig. 5. Means with standard deviation for: (a) Motivation before starting the experiment and Mann–Whitney U test results; (b) SUS score and Mann–Whitney U test results; (c) UEQ factors (pragmatic and hedonic) and all item/question together (overall) with Mann–Whitney U test results.

(KEYWORD and KEYWORD + VISUALISATION) are shown in Figure 6a. The data summarised in Figure 6a is analysed using a between-within subjects ANOVA on the 20% trimmed means [37].

A significant main effect of INTERFACE on task completion time could be detected ($F(1,60) = 14.06$, $p < 0.001$, $n^2 p = 0.19$). Here, the completion time was significantly lower for AR condition($\bar{x} = 618.57s$, $SD = 77.92s$) compared to the NON-AR condition ($\bar{x} = 698.41s$, $SD = 90.59s$). Also, a significant main effect of INSTRUCTION MODE on task completion time could be detected ($F(1,60) = 31.19$, $p < 0.001$, $n^2 p = 0.34$), such that KEYWORD + VISUALISATION ($\bar{x} = 599.04s$, $SD = 74.99s$) resulted in a significantly lower completion time than KEYWORD ($\bar{x} = 717.95s$, $SD = 93.52s$). There was no significant interaction effect found between INTERFACE and INSTRUCTION MODE ($F(1,60) = 0.25$, $p > 0.05$, $n^2 p < 0.001$).

5.8 Learning Efficiency

The average learning efficiencies for immediate recall and delayed recall across study conditions are shown in Figure 6a and Figure 6b respectively. For the definition of learning efficiency refer to Sect. 4.5. The data summarised in Figure 6a and Figure 6b are analysed using a between-within subjects ANOVA on the 20% trimmed means [37].

Statistical analysis in Figure 6b and Figure 6c showed no significant effect of INTERFACE for participants' learning efficiency for immediate recall ($F(1,60) = 9e-6$, $p > 0.05$, $n^2 p < 0.001$) or delayed recall ($F(1,60) = 9e-5$, $p > 0.05$, $n^2 p < 0.001$). A significant main effect of INSTRUCTION MODE on participants' learning efficiency for immediate recall could be detected ($F(1,60) = 34.14$, $p < 0.001$, $n^2 p = 0.36$). There, the learning efficiency was significantly higher in KEYWORD + VISUALISATION support ($\bar{x} = 0.92$, $SD = 0.23$) compared to the KEYWORD ($\bar{x} = -0.92$, $SD = 0.23$). A significant main effect on participants' learning efficiency for delayed recall also could be detected ($F(1,60) = 41.25$, $p < 0.001$, $n^2 p = 0.41$). There, the learning efficiency was significantly higher in KEYWORD + VISUALISATION instruction mode ($\bar{x} = 1.07$, $SD = 1.15$) compared to the KEYWORD ($\bar{x} = -1.07$, $SD = 1.32$) mode. There was no significant interaction effect found between INTERFACE and INSTRUCTION MODE for in immediate recall ($F(1,60) = 0.07$, $p > 0.05$, $n^2 p < 0.001$) or delayed recall ($F(1,60) = 0.18$, $p > 0.05$, $n^2 p < 0.001$).

6 Discussion and Future Directions

For this study, we developed an AR system called VocabulARy, that supports learning new Japanese words, but can be expanded to support other languages. The system was used to evaluate user experience, system usability, mental effort, motivation and memory recall when shown keywords over the objects vs. keywords together with a visualisation of the objects. These were compared in two interfaces: AR (Microsoft HoloLens 2) and a NON-AR (Android tablet computer). We used the two interfaces to investigate whether showing keywords and visualisations in context of immediate surrounding compared to the context provided on the virtual scene on the screen results in any performance difference.

6.1 Usability and User Experience

The results of the study show that participants evaluated both AR and NON-AR prototypes with good usability scores, clearly higher than average (68) and no significant difference between the two could be found. In addition, during the study we did not observe any readability problems (e.g. none of the users tried to zoom in on the tablet computer in order to make it easier to view presented information and none of the AR HMD users were observed to move very close to augmentations). This provided a good basis for further investigation, as we wanted to make the comparison as fair as possible by trying to not influence learning performance with usability issues as well as by making both conditions as comparable as possible (see Sect. 3.1).

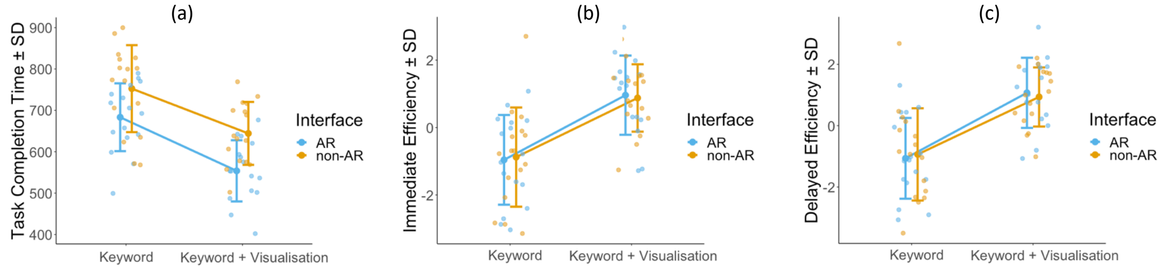
It has been shown for example that the unfamiliarity with AR could result in lower performance as has been reported in prior work [78]. In addition, the user experience in both conditions has been rated very positively with a higher score for AR regarding the hedonic factors represented by Stimulation and Novelty. This makes sense, as AR is still an exciting and less widely used technology compared to tablet computers for many users.

A recent study also revealed that the size of images affects our ability to remember image content during naturalistic exploration [38]. Although our study did not involve naturalistic exploration, it clearly steered participants' attention, and we made extra care that both AR and NON-AR showed comparable imagery, it would still be interesting to explore if the size of imagery has an effect on the ability to memorise vocabulary words.

The prolonged use of HMDs in the current form factor can also influence the usability and thus performance of users. Some studies have already investigated the effects of the HMDs weight, their pressure on the face, latency, image quality and the authenticity of the representation of digital objects [20, 21, 69, 72]. However, these issues will likely be addressed with future development of HMDs.

6.2 AR vs Non-AR

Our results show that immediate recall (a recall of words right after the study) in the AR system is significantly higher compared to the NON-AR system. However, no statistical significance was detected for delayed recall (a recall of the same words a week after the study). Nonetheless, it is important to note that significance was only marginally missed for delayed recall. The results can thus not confirm the outcomes of a previous study conducted by Ibrahim et al. [28] who report a significantly better performance of the AR system compared to FLASHCARDS for both immediate and delayed recall. One of the reasons, and also



Task Completion Time	SS	df	MS	F	p	n²p
Interface	1e+5	1	1e+5	14.06	0.00 ***	0.19
Instruction Mode	2e+5	1	2e+5	31.19	0.00 ***	0.34
Interface * Instruction Mode	1811	1	0	0.25	0.62	0.00

Immediate Efficiency	SS	df	MS	F	p	n²p
Interface	1e-5	1	1e-5	9e-6	0.99	0.00
Instruction Mode	53.93	1	53.93	34.14	0.00 ***	0.36
Interface * Instruction Mode	0.11	1	0.11	0.07	0.79	0.00

Delayed Efficiency	SS	df	MS	F	p	n²p
Interface	1e-4	1	1e-4	9e-5	0.99	0.00
Instruction Mode	64.37	1	64.37	41.25	0.00 ***	0.41
Interface * Instruction Mode	0.28	1	0.28	0.18	0.68	0.00

Fig. 6. Means with standard deviation and ANOVA results for: (a) Task-completion-time in seconds; (b-c) immediate recall and delayed recall learning efficiency.

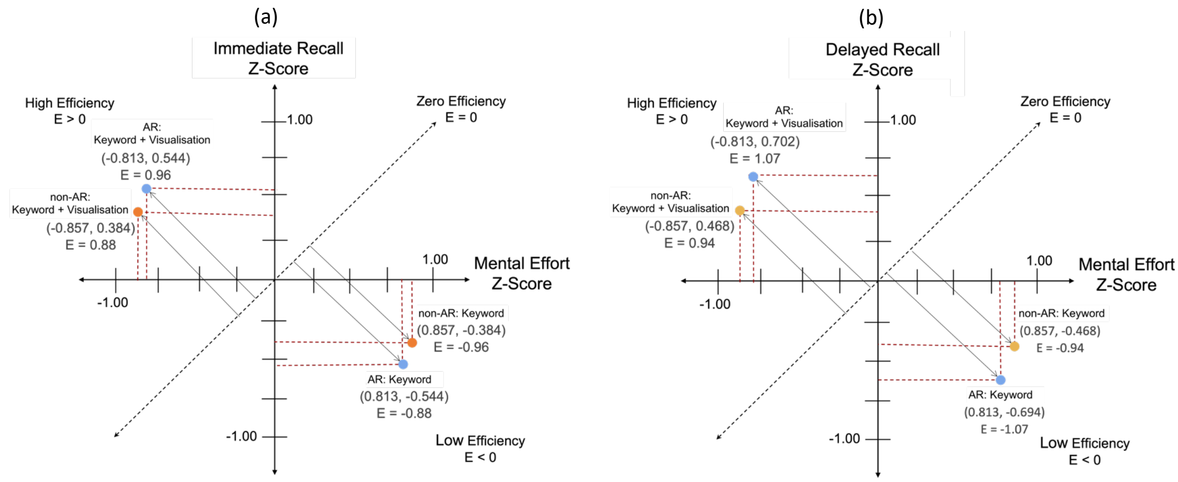


Fig. 7. Learning efficiency: (a) immediate recall learning efficiency; (b) delayed recall learning efficiency.

a major difference between this and our study, might be that in the aforementioned work, the learning methods were not identical in both conditions, which could have placed the AR system at an advantage. For example, with flashcards the word was shown on the opposite side to the image depicting word meaning; thus, the image and word were never shown together. This was not the case in our AR condition where the word annotations were always visible for a selected object in the scene. In comparison, we carefully designed our experiment to minimise any such confounding variable that might influence the results.

Furthermore, participants expressed a significantly higher level of motivation in the AR condition. This is in line with previous work [12, 34] and should be considered when interpreting our results since motivation can be an important factor in learning [8] and technology can play a significant role in this [35]. What causes higher motivation falls out of the scope of this study; however, one could hypothesise that the novelty of the AR plays an important role. The observations show that participants were excited about testing the AR HMD compared to using a tablet. This introduces a need for a longitudinal study of using AR in vocabulary learning, as the influence of motivation might decrease with increasing familiarity with the system.

Interestingly, the AR condition also outperformed the NON-AR condition in task completion time (about 11% faster). This results show that participants were able to learn all words faster in AR condition compared to the NON-AR condition whilst also achieving higher immediate recall scores. This result is also somewhat surprising as AR is at a disadvantage to NON-AR for activating objects. That is, target selection in our setup was typically faster on a tablet computer compared to in mid-air tapping on a HMD. Furthermore, a tablet computer also offered instant access to all buttons at the same distance, whereas users need to physically move to activate some of the buttons in AR. One future direction could involve making users spend the same amount of time in both conditions and explore if this would further improve the performance of AR condition.

6.3 Keyword vs. Keyword + Visualisation

Regarding our second independent variable INSTRUCTION MODE, our results clearly show that vocabulary learning can be improved beyond the traditional keyword method, by augmenting the keywords with animated 3D visualisations. We found overwhelming evidence for this in all metrics, such as immediate and delayed recall, learning efficiency, mental effort, and even task completion time.

This is in line with observations by Shapiro and Waters [67], who reported that the level of visual imagery of a word enhances vocabulary learning. In this work, we go beyond simple imagery and show that the 3D animated content is potentially an even more effective approach. This opens up another direction for future work involving the necessity for detailed comparison of the effect of different visualisation techniques.

As mentioned, our results showed that providing visualisations for keywords reduced the mental effort for vocabulary learning. However, reducing mental effort in learning scenarios can also result in reduced learning outcomes. For example, Salmon showed that the amount of invested mental effort positively correlates with learning efficiency [61]. Knowing this, we could expect a decline in performance of immediate and delayed recall. One reason why this did not happen in our case might be the fact that enough effort was needed in order to complete the task (moving, tapping, remembering). One way to increase the mental effort would be to require users to come up with their own associations for keywords instead of providing predefined keywords as in our study. Providing predefined keywords might not be in line with user's mental model, thus making it difficult for the user to (mentally) visualise them. This could have made visualisations in our study more important. However, previous research suggests that users might have difficulties coming up with their own keywords and that predefined keywords lead to better learning outcomes [4]. Despite, future studies should explore if the difference persists also when personalised keywords are used in learning scenario presented in this study.

6.4 Implications and Design Recommendations

The benefits of the keyword method over other learning techniques are well known [30, 51]. We have shown that the keyword method can provide even better results by adding animated visualisations that depict the keyword itself. This is an important implication for designing such applications for vocabulary learning. However, this also opens up several questions. For example, do animations of visualisations of keywords contribute to the learning outcome or would visualisation without an animation result in comparable efficiency?

One of the most important things to consider in designing such a system is the keywords or words from the language a learner speaks that sound similar to the word being learnt. In our prototype, we only used a limited set of vocabulary for which we were able to find appropriate keywords and accompanying visualisations. Finding these keywords and visualisations takes time, which needs to be considered when thinking about applying this method in practice. And it is no necessary that every word would have an appropriate keyword. Crowd-sourcing could be one approach to tackle this problem. Additionally, approaches for automating the process of finding keywords already exist [2]. Also, our future direction will involve investigating the effect of asking users to choose their own keywords and visualisations. A system that would use a combination of these approaches could probably satisfy a variety of learning types.

Another thing to keep in mind, and we are not aware of any study investigating it, is the fact that the vast number of keywords might be overwhelming for users. One of the unanswered question is thus how many keywords is recommended to provide at one time (in our study only one was shown at the time to direct users).

For the purpose of this study we used marker based tracking to initialise the settings in AR. As such, our system was linked to a particular physical space. To enable wider adoption, another space-independent object recognition technique should be used as discussed in Sect. 3.2. Such a system would also need to have a database of objects with the corresponding keywords available upfront so when users look at a physical scene AR visualisations would be fetched on the fly.

Only two participants used all the available time to learn and all 10 words were correctly remembered in 23% of immediate and 4% of delayed recall tests. For immediate recall we might have reached the ceiling effect, and making the task more difficult would highlight even greater changes between the test conditions. This was even more obvious for delayed recall, where only very few users finished the test with no errors. This could be taken in consideration when building such a system – during the testing phase the system should try to increase the level of engagement and encourage users to take more time, while and after testing the system should encourage users to rethink about wrong answers.

6.5 Limitations

As explained in Sect. 5 gender did not have a significant effect on the results of the study. However, future work should look into a possible gender bias in more detail with a higher number of participants, as our result on this is not conclusive.

Another thing to consider in our study is age bias. However, the age group studied is highly mobile, spending extended periods of time in foreign countries (for example, the EU Erasmus+ programme alone funds more than half a million exchanges yearly [16]). As such, this group could benefit from an improved vocabulary learning system. Nevertheless, the results cannot be generalised over the whole population, and expanding the study to other age groups and exploring the effect of age on the proposed learning system is an important future direction.

Further, we only used nouns in our prototype. More specifically, all nouns were associated with objects. In fact, a number of studies have shown that concrete terms (e.g., nouns such as bread) are better remembered than abstract terms (e.g., abstract nouns and verbs) [70]. The benefits of in-situ learning with AR will therefore be reduced when abstract terms are considered as it becomes difficult to make them relevant to context of users' immediate environment. Nevertheless, future studies could focus on exploring the potential of 3D AR animation to make abstract terms visually more accessible.

As mentioned in the paper, our prototype was only tested for a short time on a limited vocabulary. To further validate our findings, the vocabulary should be expanded and tested over a longer period of time. Especially the effect of higher motivation in AR could wear off as users become more familiar with the system. Additionally, we only measured the immediate recall (immediately after participants had completed the task) and a delayed recall (a week after participants had completed the task) of the vocabulary learned. Future work should also consider recall after longer periods of several weeks. This could also be combined with repeating the learning phase in certain intervals, as it is normally done when learning vocabulary.

7 Conclusion

Learning vocabulary can be enhanced when encountering words in context. This context can be afforded by the place or activity people are engaged with. For this purpose we developed VocabulARy, a HMD AR system that visually annotates objects in the user's surroundings, with the corresponding English (first language) and Japanese (second language) words to enhance the language learning process. In addition to the written and audio description of each word, we also present the user with a keyword and its animated 3D visualisation to enhance memory retention.

We evaluated our prototype by comparing it to an alternate AR system that does not show any additional visualisation of the keyword, and also, we compare it to two non-AR systems on a tablet, one with and one without visualising the keyword. Our results indicate that AR outperforms the NON-AR system regarding short-term retention, mental effort and task-completion time. Additionally, the visualisation approach scored significantly higher than only showing the written keyword with respect to immediate and delayed recall, learning efficiency, mental effort and task-completion time. Visualisation of keywords thus proved more efficient compared with the traditional keyword method only and opens new avenues for future improvements in AR enabled vocabulary learning systems.

Acknowledgments

The authors wish to thank Cuauthli Campos for helping with preparing the video and all the volunteers who participated in the user study.

This research was supported by European Commission through the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund). We also acknowledge support from the Slovenian research agency ARRS (program no. BI-DE/20-21-002, P1-0383, J1-9186, J1-1715, J5-1796, and J1-1692).

References

[1] M. Amiryousefi and S. Ketabi. Mnemonic instruction: A way to boost vocabulary learning and recall. *Journal of Language Teaching & Research*, 2(1), 2011.

[2] O. Anonthanasap, C. He, K. Takashima, T. Leelanupab, and Y. Kitamura. Mnemonic-based interactive interface for second-language vocabulary learning. *Proceedings of the Human Interface Society, HIS*, 14, 2014.

[3] D. Ary, L. C. Jacobs, C. K. S. Irvine, and D. Walker. *Introduction to research in education*. Cengage Learning, 2018.

[4] R. C. Atkinson. Mnemotechnics in second-language learning. *American psychologist*, 30(8):821, 1975.

[5] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[6] H. D. Brown and S. T. Gonzo. *Readings on second language acquisition*. Allyn & Bacon, 1995.

[7] R. O. Castle and D. W. Murray. Object recognition and localization while tracking and mapping. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pp. 179–180, 2009. doi: 10.1109/ISMAR.2009.5336477

[8] Y.-C. Chen and P.-C. Chen. The effect of english popular songs on learning motivation and learning performance. *WHAMPOA-An Interdisciplinary Journal*, 56:13–28, 2009.

[9] R. C. Clark, F. Nguyen, and J. Sweller. *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. John Wiley & Sons, 2011.

[10] A. D. Cohen and E. Aphek. Retention of second-language vocabulary overtime: Investigating the role of mnemonic associations. *System*, 8(3):221–235, 1980.

[11] J. Cohen. Statistical power analysis for the behavioral sciences. *England: Routledge*, 1988.

[12] C. S. C. Dalim, M. S. Sunar, A. Dey, and M. Billinghurst. Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, 134:44–64, 2020.

[13] F. Draxler, A. Labrie, A. Schmidt, and L. L. Chuang. Augmented reality to enable users in learning case grammar from their real-world interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[14] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.

[15] D. Edge, E. Searle, K. Chiu, J. Zhao, and J. A. Landay. Micromandarin: mobile language learning in context. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 3169–3178, 2011.

[16] Erasmus+. Factsheets and statistics on Erasmus+. `https://erasmus-plus.ec.europa.eu/node/2585`. 2022-05-24.

[17] P. A. Freund, J.-T. Kuhn, and H. Holling. Measuring current achievement motivation with the qcm: Short form development and investigation of measurement invariance. *Personality and Individual Differences*, 51(5):629–634, 2011.

[18] Y. Fujimoto, G. Yamamoto, H. Kato, and J. Miyazaki. Relation between location of information displayed by augmented reality and user's memorization. In *Proceedings of the 3rd Augmented Human International Conference*, pp. 1–8, 2012.

[19] T. Georgiev, E. Georgieva, and A. Smrikarov. M-learning-a new stage of e-learning. In *International conference on computer systems and technologies-CompSysTech*, vol. 4, pp. 1–4, 200.

[20] J. Guo, D. Weng, Z. Zhang, H. Jiang, Y. Liu, Y. Wang, and H. B.-L. Duh. Mixed reality office system based on maslow's hierarchy of needs: Towards the long-term immersion in virtual environments. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 224–235. IEEE, 2019.

[21] J. Guo, D. Weng, Z. Zhang, Y. Liu, and Y. Wang. Evaluation of maslows hierarchy of needs on long-term use of hmds–a case study of office environment. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 948–949. IEEE, 2019.

[22] A. K. Halabi. Applying an instructional learning efficiency model to determine the most efficient feedback for teaching introductory accounting. *Global Perspectives on Accounting Education*, 3(1):6, 2006.

[23] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[24] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[25] A. Hautasaari, T. Hamada, K. Ishiyama, and S. Fukushima. Vocabura: A method for supporting second language vocabulary learning while walking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–23, 2019.

[26] A. Hinderks, M. Schrepp, and J. Thomaschewski. A benchmark for the short version of the user experience questionnaire. In *WEBIST*, pp. 373–377, 2018.

[27] M. B. Ibáñez, Á. Di Serio, D. Villarán, and C. D. Kloos. Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, 71:1–13, 2014.

[28] A. Ibrahim, B. Huynh, J. Downey, T. Höllerer, D. Chun, and J. O'donovan. Arbis pictus: A study of vocabulary learning with augmented reality. *IEEE transactions on visualization and computer graphics*, 24(11):2867–2874, 2018.

[29] B. Khoshnevisan and S. Park. Affordances and pedagogical implications of augmented reality (ar)-integrated language learning. In *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education*, pp. 242–261. IGI Global, 2021.

[30] M. E. King-Sears, C. D. Mercer, and P. T. Sindelar. Toward independence with keyword mnemonics: A strategy for science vocabulary instruction. *Remedial and Special Education*, 13(5):22–33, 1992.

[31] G. F. Kuder and M. W. Richardson. The theory of the estimation of test reliability. *Psychometrika*, 2(3):151–160, 1937.

[32] S. Kumar Basak, M. Wotto, and P. Belanger. E-learning, m-learning and d-learning: Conceptual definition and comparative analysis. *E-learning and Digital Media*, 15(4):191–216, 2018.

[33] J. R. Lewis and J. Sauro. The factor structure of the system usability scale. In *International conference on human centered design*, pp. 94–103. Springer, 2009.

[34] S. Li, Y. Chen, D. M. Whittinghill, and M. Vorvoreanu. A pilot study exploring augmented reality to increase motivation of chinese college students learning english. In *2014 ASEE Annual Conference & Exposition*, pp. 24–85, 2014.

[35] M.-H. Lin, H.-g. Chen, et al. A study of the effects of digital learning on learning motivation and learning outcome. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7):3553–3564, 2017.

[36] S. E. C. LTD. Samsung galaxy - the official samsung galaxy site. `https://developer.vuforia.com/`. Accessed: 2022-07-27.

[37] P. Mair and R. Wilcox. Robust statistical methods in r using the wrs2 package. *Behavior research methods*, 52(2):464–488, 2020.

[38] S. Masarwa, O. Kreichman, and S. Gilaie-Dotan. Larger images are better remembered during naturalistic encoding. *Proceedings of the National Academy of Sciences*, 119(4):e2119614119, 2022. doi: 10.1073/pnas.2119614119

[39] M. A. Mastropieri and T. E. Scruggs. *Teaching students ways to remember: Strategies for learning mnemonically.* Brookline Books, 1991.

[40] T. Mayes and S. De Freitas. Learning and e-learning. *Rethinking pedagogy for a digital age*, pp. 13–25, 2007.

[41] C. McLoughlin and M. J. Lee. Personalised and self regulated learning in the web 2.0 era: International exemplars of innovative pedagogy using social software. *Australasian Journal of Educational Technology*, 26(1), 2010.

[42] Microsoft. Microsoft hololens — mixed reality technology for business. `https://www.microsoft.com/en-us/hololens`. Accessed: 2022-07-27.

[43] NASA. Nasa tlx: Task load index, 2006.

[44] P. Nicholson. *A History of E-Learning*, pp. 1–11. Springer Netherlands, Dordrecht, 2007. doi: 10.1007/978-1-4020-4914-9_1

[45] D. Nunan. *Second Language Teaching & Learning.* ERIC, 1999.

[46] F. G. Paas and J. J. Van Merriënboer. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors*, 35(4):737–743, 1993.

[47] A. Paivio and A. Desrochers. Mnemonic techniques in second-language learning. *Journal of Educational Psychology*, 73(6):780, 1981.

[48] I. Pearlman. Effectiveness of keyword versus direct instruction on vocabulary acquisition by primary-grade handicapped learners. *Bulletin of the Psychonomic Society*, 28(1):14–16, 1990.

[49] 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.

[50] M. Pressley, J. R. Levin, and H. D. Delaney. The mnemonic keyword

method. *Review of Educational Research*, 52(1):61–91, 1982.

[51] M. Pressley, J. R. Levin, N. A. Kuiper, S. L. Bryant, and S. Michener. Mnemonic versus nonmnemonic vocabulary-learning strategies: Additional comparisons. *Journal of Educational Psychology*, 74(5):693, 1982.

[52] PTC. Vuforia developer portal. `https://developer.vuforia.com/`. Accessed: 2022-07-27.

[53] A. L. Putnam. Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, 1(2):130, 2015.

[54] M. R. Raugh and R. C. Atkinson. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 67(1):1, 1975.

[55] M. R. Raugh, R. D. Schupbach, and R. C. Atkinson. Teaching a large russian language vocabulary by the mnemonic keyword method. *Instructional science*, 6(3):199–221, 1977.

[56] F. Rheinberg, R. Vollmeyer, and B. D. Burns. Fam: Ein fragebogen zur erfassung aktueller motivation in lern-und leistungssituationen (langversion, 2001). *Diagnostica*, 2:57–66, 2001.

[57] F. Rheinberg, R. Vollmeyer, and B. D. Burns. Qcm: A questionnaire to assess current motivation in learning situations. *Diagnostica*, 47(2):57–66, 2001.

[58] N. Sagarra and M. Alba. The key is in the keyword: L2 vocabulary learning methods with beginning learners of spanish. *The modern language journal*, 90(2):228–243, 2006.

[59] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. Kelly, and A. J. Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1352–1359, 2013. doi: 10.1109/CVPR.2013.178

[60] M. Salman and M. J. Pearson. Whisker-ratslam applied to 6d object identification and spatial localisation. In V. Vouloutsi, J. Halloy, A. Mura, M. Mangan, N. Lepora, T. J. Prescott, and P. F. Verschure, eds., *Biomimetic and Biohybrid Systems*, pp. 403–414. Springer International Publishing, Cham, 2018.

[61] G. Salomon. Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of educational psychology*, 76(4):647, 1984.

[62] M. E. C. Santos, T. Taketomi, G. Yamamoto, M. M. T. Rodrigo, C. Sandor, H. Kato, et al. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1):1–23, 2016.

[63] M. Schrepp. Ueq-user experience questionnaire, 2019.

[64] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *IJIMAI*, 4(6):103–108, 2017.

[65] B. B. Schultz. Levene's test for relative variation. *Systematic Zoology*, 34(4):449–456, 1985.

[66] H. Schuman, S. Presser, and J. Ludwig. Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, 45(2):216–223, 1981.

[67] A. M. Shapiro and D. L. Waters. An investigation of the cognitive processes underlying the keyword method of foreign vocabulary learning. *Language teaching research*, 9(2):129–146, 2005.

[68] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[69] R. Shen, D. Weng, S. Chen, J. Guo, and H. Fang. Mental fatigue of long-term office tasks in virtual environment. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 124–127. IEEE, 2019.

[70] R. N. Shepard. Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and verbal Behavior*, 6(1):156–163, 1967.

[71] M. Spitzer. M-Learning? When it comes to learning, smartphones are a liability, not an asset. *Trends in Neuroscience and Education*, 4(4):87–91, 2015. doi: 10.1016/j.tine.2015.11.004

[72] F. Steinicke and G. Bruder. A self-experimentation report about long-term use of fully-immersive technology. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 66–69, 2014.

[73] M. P. Strzys, S. Kapp, M. Thees, P. Klein, P. Lukowicz, P. Knierim, A. Schmidt, and J. Kuhn. Physics holo. lab learning experience: using smartglasses for augmented reality labwork to foster the concepts of heat conduction. *European Journal of Physics*, 39(3):035703, 2018.

[74] G. M. Sullivan and A. R. Artino Jr. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542, 2013.

[75] U. Technologies. Unity real-time development platform. `https://unity.com/`. Accessed: 2022-07-27.

[76] C. D. Vazquez, A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2172–2179, 2017.

[77] Z. Wei. Does teaching mnemonics for vocabulary learning make a difference? putting the keyword method and the word part technique to the test. *Language Teaching Research*, 19(1):43–69, 2015.

[78] M. Wille, L. Adolph, B. Grauel, S. Wischniewski, S. Theis, and T. Alexander. Prolonged work with head mounted displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program*, pp. 221–224, 2014.

[79] H.-K. Wu, S. W.-Y. Lee, H.-Y. Chang, and J.-C. Liang. Current status, opportunities and challenges of augmented reality in education. *Computers & education*, 62:41–49, 2013.

[80] S. Yang and B. Mei. Understanding learners' use of augmented reality in language learning: insights from a case study. *Journal of Education for Teaching*, 44(4):511–513, 2018.

Chapter 7

Conclusion

Nowadays, harnessing the wonders of technology means that we can learn not only in classrooms, but anywhere, anytime in our daily lives. This factor has motivated the studies described in the previous chapters of this thesis. While technological innovations have become great "door openers" for learning, thought should also be given to the obstacles they have yet to overcome, such as the impact on established learning approaches and experiences. To understand and solve this problem, it is necessary to explore from many perspectives how technology can and could be meaningfully integrated into our daily lives. In this work, we have chosen to look more closely at the use of extended reality (XR) technology from the perspective of designing adaptive guidance and creating experiential learning (ExL) environments to explore the benefits and identify areas for future development. This concluding chapter will bring together the key findings of the research articles presented in this thesis. Section 7.1 considers how the findings of different user studies address the research questions presented in Section 1.2 followed by the recap of research contributions of this thesis (Section 7.2). This is followed by a discussion regarding the limitations of this research and possible directions for future work in Section 7.3. Lastly, this chapter closes in Section 7.4 with concluding remarks around the value of this work.

7.1 Experiential Learning in XR

This work investigated the *potential of XR technologies in designing and implementing adaptive learning environments that aim to tailor the learning experience to meet each learner's unique needs, enhance their understanding and increase engagement through interactive and stimulating environments supported by effective guidance.* It explored the effectiveness of different configurations of XR, ExL, and instructional guidance in creating immersive learning experiences.

We structure the research presented around four research questions (RQs) each with several corresponding hypothesis (see Section 1.2). To answer the research questions, three studies were conducted – each focusing on an experiential learning system developed based on a particular configuration of the design space presented in Chapter 3 (i.e., focusing on different configurations of XR based instruction to support experiential learning). Each study resulted in a different research publication, as presented in Chapter 4, Chapter 5, and Chapter 6.

In Study 1 (*VRNav Study:* Chapter 4), our objective was to explore how different *Display Configurations* and *Data Modalities* would affect user engagement and understanding of space with minimal guidance. We observed and evaluated different representation methods –i.e., VR, desktop and 2D paper-based floor plans – on spatial learning in ExL environments, through various data collection and standard questionnaires combining qualitative and quantitative research methods. The findings from this study provided valuable insights into the impact of immersive VR and other representation methods on user experience, user engagement dimensions, and spatial understanding of ExL environments. These findings helped to address RQ1 and RQ4 and are discussed in detail within the following subsections (7.1.1 and 7.1.2).

In Study 2 (*Arigatō Study*: Chapter 5), we aimed to investigate the influence of the *Experiential Learning* (ExL) cycle and different adaptation factors of *Guidance* –i.e., the amount and the type of the guidance – on user experience, engagement, and learning outcomes in a language learning task. This was achieved by designing a HMD based AR language learning system that offers immediate assistance to guide learners through different stages of the ExL cycle addressing RQ1. By examining the amount and type of guidance provided, we addressed RQ3. The findings from this study also helped identify XR interventions that could be used to enhance instructional guidance addressing RQ2. Furthermore, this study investigated how the dynamic adaptations of guidance influenced user experience, engagement, and performance in a language learning task within an immersive AR ExL environment, which contributed to addressing RQ4.

In Study 3 (*VocabulARy Study*: Chapter 6), we focused on investigating the impact of different *Display Configurations* (i.e., AR HMD and non-AR 2D display) and different types of *Guidance* (i.e., keyword only and keyword + visualisations) on user engagement and learning performance in a vocabulary learning task. This was achieved by designing an HMD application that visually annotated objects in the user's surroundings and compared it to a non-AR system on a tablet. This study directly addressed RQ1 and RQ4, and provided valuable findings into the importance and effectiveness of AR and different types of guidance in immersive vocabulary learning environments. These are discussed in detail within the following subsections.

While each study contributes to the overall research aim, it is crucial to merge the findings together to gain clearer insights into how the four research questions were addressed. By doing so, we can enhance our understanding of the impact of different configurations of XR, ExL, and instructional guidance on user experience, engagement and learning performance in ExL environments.

7.1.1 Impact on User Experience and Engagement

The impact of XR technology on the experience and engagement of the users was an underlying theme of each of the studies conducted in this thesis.

Engagement can be categorised into affective, cognitive, and behavioural engagement [12]. In Study 1 (*VRNav Study*), presented in Chapter 4, we hypothesised that the different *Display Configurations* and *Data Modalities* –i.e., representation methods – would affect engagement (H1-2, H4-1). The results of VRNav Study clearly confirm these two hypotheses. For example, we measured behavioural engagement for VR and desktop conditions by observing where users spent their time and how much time they spent in different locations and found out that people spent nearly 40% more time moving in VR compared to the desktop. Perhaps this is the case because moving in space has been pointed out as one of the pivotal affordances in VR [143, 54]. This shows that the display configuration has a clear effect on engagement. On the other hand, in paper condition we could also observe users being more interested in renders (egocentric view) than in floor plans (allocentric view). This was expected as floor plans can be hard to mentally translate to an egocentric view for non professionals, which has also been found in this study. This shows the data modalities have a clear effect on engagement.

In addition, in VRNav Study, we also measured affective engagement through analysis of facial expression for desktop and paper conditions. The amount of positive feelings was higher in the desktop condition. Significantly more negative feelings were recorded in the paper condition, while these were almost absent in the desktop condition. Affective engagement is directly linked to the sense of presence and emotions. For example, experience in anxious or relaxing VR environments increased these emotions as well as the sense of presence [106]. Studies on learning in VR also report on higher emotional arousal compared to content presented on a computer monitor [97]. We could not measure affective engagement in the VR condition; although, since the virtual environment was the same in the desktop condition as in the VR condition, and since users spent considerably more time in the VR condition, we can assume that the VR condition would show even higher affective engagement based on the increased behavioural engagement. Again this shows the display configuration affects engagement.

In Study 2 (*Arigatō Study*), presented in Chapter 5, we hypothesised that different adaptation factors of *Guidance* –i.e., the type and the amount of guidance – would affect engagement and enable the creation of more effective immersive adaptive educational systems (H4-1, H3-1). The results of Arigatō Study confirm these hypothesise. For example in Arigatō Study, we measured behavioural engagement for fixed-guidance and adaptive-guidance by observing how participants interacted with the AR language learning system and how much time they spent in each condition. We could observe participants being more interested in adaptive-association type of guidance than the fixed-association type of guidance, even though they completed the task in less time. This was expected, since the keyword method has been suggested to motivate learners to be more creative and to use their minds more productively [105]. This clearly indicates the the amount and type of guidance affect engagement, which in turn enable the creation of more effective immersive adaptive educational systems.

In addition, the Arigatō Study also measured the cognitive engagement for study conditions by analysing how much mental effort participants invested in completing the task. We found that the mental effort with the adaptive-amount of guidance was significantly lower compared to the fixed-amount of guidance. This is in line with the literature, as prior work has shown that repeating the same information could increase the amount of cognitive work and consequently the mental effort [109]. We further identified that the mental effort in the adaptive-association type of guidance was lower than in fixed-association type of guidance. This was expected, as the keyword method connects new learning to prior knowledge through the use of visual and/or acoustic cues [101, 84], which could lower the mental effort. This again confirms that the amount and type of guidance affects engagement.

The ability to change the amount of guidance dynamically in the Arigatō Study clearly shows it is possible to remodel the guidance in XR environments based on the engagement and learning outcomes (H2-1). The study also revealed indirect evidence that the explicit guidance can support the learner to move from one stage to the other in Kolb's ExL cycle (H2-2).This finding is based on good learning outcomes and above average user experience scores reported by UEQ questionnaire. Both would not be possible if learners would not successfully move through all stages of Kolb's ExL cycle [69].

In Study 3 (*VocabulARy Study*), presented in Chapter 6, we hypothesised that different *Display Configurations* (i.e., AR HMD and non-AR 2D display) and different types of *Guidance* (i.e., keyword only and keyword + visualisations) would affect engagement (H4-1, H4-2). The results of VocabulARy Study confirm these hypotheses. For example, in VocabulARy Study, we measured behavioural engagement by observing how participants interacted with the vocabulary learning system and how much time they spent in each condition. We could observe participants being more interested in learning in AR compared to the non-AR condition, even though they completed the task in less time. This confirms that the display configuration has an effect on engagement. In Study 3 (*VocabulARy Study*), we also measured the cognitive engagement for study conditions by analysing how much mental effort participants invested in completing the task. We found that the mental effort within the AR was significantly lower compared to the non-AR. We further identified that the mental effort in the keyword + visualisations type of guidance was lower than in keyword only type of guidance. Inline with the findings from Study 2, this was expected, as the level of visual imagery of a word enhances vocabulary learning [116]. This confirms that the type of guidance has an effect on engagement.

In all the studies we used the UEQ questionnaire which provides a benchmark to compare user experience between different systems [53]. It measures pragmatic factors of a system (including efficiency, perspicuity and dependability) and hedonic factors (including stimulation and novelty). The results showed very positive scores for all XR systems in particular for the hedonic factors. This confirms our hypothesis that XR learning environments can outperform standard learning systems (H3-1).

7.1.2 Impact on Learning Outcomes

The studies conducted as part of this thesis, provide insights into how learning outcomes are affected by the XR technology and instructions design. While Study 1 provides a 'broad & general' view of the role of learning and guidance, the findings from Study 2 and Study 3 provide a very 'concentrated & specific' perspective of the impact XR technology on learning and designing guidance.

It has long been argued that spatial understanding is the VR's most intuitive benefit [23]. In the physical world we exploit several depth cues such as stereopsis, motion parallax, perspective, and occlusion for reconstructing 3D scenes. Similar depth cues can be provided by a detailed 3D VR environment and a high level of immersion can lead to greater spatial understanding. In Study 1 (*VRNav Study*), we hypothesised that the immersive VR condition will improve learning outcomes, such as for example spatial understanding (H4-1). The results of VRNav Study confirm this hypothesis. For example we explored participants' understanding of the space by measuring participants' abilities to estimate distance and direction within the functional space in three conditions (i.e., VR, desktop and 2D paper-based floor plans). The findings from Study 1 show that users in desktop and VR conditions underestimated distances and users in VR condition were best in distance estimation with the error rate of around 7% (desktop 13%). The highly-detailed virtual environment offering several depth cues probably lead to the results comparable to real world distance estimations [140, 123]. Active exploration of the virtual environment can also lead to accurate egocentric distance between virtual objects [123].

In the study, participants also had to estimate the direction to a certain interesting area (landmark) from their egocentric position using traditional direction circles [134]. As in distance estimations, this was done for when egocentric position and users were on the same floor, when there was one floor difference (between the user and the interesting area) and two floor difference. The direction estimations in all study conditions resulted in high error rates (VR error was minimum) and we found no significant difference in direction estimation in all three study conditions. The findings in the literature show approximately 85% correctness in direction estimation while observing simple objects in non-realistic desktop based virtual environment [5]. Other studies showed that estimating directions in real, high-detailed desktop based and VR based environment are comparable and amount to around 8° error for adults [134, 59]. While estimations using direction circles had an error twice as high [134]. It has to be stresses out that in all these studies participants had clear tasks to perform before being tested. The free-style tour used in our experiment clearly affected our results. Nevertheless, these findings confirms our hypothesis that immersive VR environment improves spatial understanding.

In Study 2 (*Arigatō Study*), we hypothesised the the type and the amount of instructional guidance generated by XR environments would affect the learning outcomes whilst adaptive amount of guidance will also outperform fixed amount (H4-2, H4-3). The results of Arigatō Study confirm only the hypothesis H4-2. For example, in Arigatō Study we measured learning performance through analysis of recall (immediate recall after the study and delayed recall test 7 days after the study) and learning efficiency. Interestingly, we found that compared to the adaptive-amount of guidance, the fixed-amount of guidance group scored better in the immediate and delayed recall tests, although they have invested a significantly higher mental effort during the task. In addition, we found that the adaptive-associations type of guidance outperformed the fixed-associations type of guidance, in the immediate, delayed recall tests, and learning efficiency, while investing a significantly lower mental effort and spending less amount of time to complete the task. These finding suggest avenues for future work in archetypal user models, which might allow users to slowly adapt pre-developed personal models from a standard set of ideal users. What we would refer to as the "follow the leader" learning approach. These findings clearly confirm our hypothesis that the type and the amount of instructional guidance affects the learning outcomes, however rejects the hypothesis that the adaptive-amount is better than fixed-amount.

In Study 3 (*VocabulARy Study*), we hypothesised that the display configuration and the type of instructional guidance generated by XR environments would affect the learning outcomes and enable immersive educational systems that would outperform standard learning systems (H4-1, H4-2 and H3-1). The results of VocabulARy Study confirm these hypotheses. For example, in VocabulARy we again measured learning immediate recall, delayed recall and learning efficiency metrics to measure the learning performance. Our findings show that immediate recall (a recall of words right after the study) in the AR system is significantly higher compared to the non-AR system. However, no statistical significance was detected for the delayed recall (a recall of the same words a week after the study). Nonetheless, it is important to note that significance was only marginally missed for the delayed recall. Furthermore, participants expressed a significantly higher level of motivation in the AR condition. This is in line with previous work [37, 82] and should be considered when interpreting our results, since the motivation can be an important factor in learning [31] and technology can play a significant role as well [83]. These findings clearly confirm our hypotheses that the display configuration and the type of instructional guidance affect the learning outcomes.

The findings form Study 3 also show that the vocabulary learning can be improved beyond the traditional keyword method, by augmenting the keywords with animated 3D visualisations. We found overwhelming evidence for this in all metrics, such as immediate and delayed recall and learning efficiency. This is in line with observations by Shapiro and Waters [116], who reported that the level of visual imagery of a word enhances vocabulary learning. In this work, we go beyond simple imagery and show that the 3D animated content is potentially an even more effective approach. Our findings also show that providing visualisations for keywords reduced the mental effort for vocabulary learning. However, reducing mental effort in learning scenarios can also result in reduced learning outcomes. For example, Salmon showed that the amount of invested mental effort positively correlates with learning efficiency [110].

Knowing this, we could expect a decline in performance of immediate and delayed recall. One reason why this did not happen in our case might be the fact that enough effort was needed in order to complete the task (moving, tapping, remembering). One way to increase the mental effort would be to require users to come up with their own associations for keywords instead of providing predefined keywords as in our study. Providing predefined keywords might not be in line with user's mental model, thus making it difficult for the user to (mentally) visualise them. This could have made visualisations in our study more important. However, previous research suggests that users might have difficulties coming up with their own keywords and that predefined keywords lead to better learning outcomes [14]. Overall, it is Study 2 (*Arigatō Study*) and Study 3 (*VocabulARy* Study), which provide direct insights into the potential impact of XR technology for designing interactive, adaptive ExL environments for better learning outcomes. This clearly confirms that the XR technology has the potential to generate immersive educational systems that could outperform standard learning systems.

7.2 Contributions Recapped

This work explored the role of XR technology for designing interactive ExL systems: its impacts on learning and learning experience. Taken together, these findings contribute to the design and development of interactive XR based experiential learning systems that can meet the diverse learning needs and preferences of individual learners, leading to improved learning outcomes. As part of this, the potential for technology to be a source of disruption is acknowledged; negatively affecting the activities in the learning experiences of the learners. The findings of Chapter 4, 5 and 6 shed light on several significant contributions that are of interest to educators, researchers, and designers alike.

First, this thesis presents a range of user studies conducted to demonstrate the effectiveness of XR in designing interactive ExL environments. These studies, detailed in Chapter 4, Chapter 5 and Chapter 6, showcase how XR technology maximises user experience and engagement, consequently enhancing learning performance. Furthermore, the studies highlights successful examples of XR interventions that effectively guide learners through various stages of the ExL cycle. Chapter 4 outlines how XR can create engaging experiences while minimising the need for explicit guidance, thus fostering user engagement.

The research investigates design considerations for immersive ExL environments, focusing on adapting the type and amount of guidance provided to maximise user engagement and learning performance. Chapter 5 explores how adjusting the level of guidance influences user engagement, while Chapter 4 delves into adapting the type of guidance to enhance user engagement and learning outcomes.

A comprehensive analysis of various variables, including user experience, dimensions of user engagement and learning outcomes, is presented in Chapter 7. This analysis aims to better understand the usability of XR for designing interactive and adaptive ExL environments, consolidating insights gained from empirical research.

In summary, this thesis provides educators, researchers, and designers with evidence of XR's usability in designing interactive ExL environments, successful examples of XR interventions for guidance and support, design considerations for maximising user engagement and learning performance, and comprehensive analyses of user engagement and learning outcomes. The findings contribute to the advancement of XR technology in education and lay the foundation for the development of innovative interactive and experiential learning systems in the future.

7.3 Limitations and Future Work

As with many research projects there are limitations which should be acknowledged. As described in Section 1.3, the mixed methodological approach to the study design, data collection and data analysis itself provides limitations to the results presented in the studies. More specifically, failings in the experimental set-up of a study offers an opportunity to critique the findings presented. For example, if the sample size is too small or not representative of the population of interest, the results may not be statistically significant or applicable to a larger population. Similarly, if the control group is not properly defined or not properly controlled, it may lead to confounding variables that can skew the results. Additionally, as with all studies, the possibilities for bias (either from the researcher, the data collection, the data analysis or from participants), should all be considered as sources of potential error. For example using questionnaires for measuring certain aspects such as motivation, system usability, user experience and mental effort can result in social desirability bias [36] as users could answer them to please the researcher.

The average age of the participants of the studies within this thesis was 27 years (mostly university students except for some participants of Study 01), which presents a possible age bias. Choosing participants of this age group was a decision based on several factors: access to the participants; the age group studied is highly mobile; and ability to articulate thoughts and opinions. However, this has meant that the work has not addressed younger children or older participants, or those with cognitive impairments. Furthermore, the cultural background of our participants were comparable. As a result, our findings cannot be said to be indicative, representative or generalised over the whole population, in every region of the world. Thus expanding the studies to other age groups, cultural backgrounds and exploring the effects on the proposed systems is an important future direction.

Furthermore, our systems were only tested for a shorter period of time. To further validate our findings, the systems should be expanded and tested over a longer period of time. Especially, the effect of higher motivation and engagement in XR environments could wear off as users become more familiar with the system. Additionally, learning performances of Study 2 (*Arigatō Study*: Chapter 5) and 3 (*vocablARy Study*: Chapter 6) were measured only, immediately after participants had completed the task and a week after participants had completed the task. Future work should also consider measure performances after longer periods of several weeks. This could also be combined with repeating the learning phase in certain intervals.

Finally, the technology that was evaluated in this work is itself a limitation. The results presented in this work have limited applicability to the technologies used within the studies. Furthermore, as technology continues to develop and improve, some of the usability or experiential issues identified may now have changed or no longer exist.

7.4 Concluding Remarks

As we move into the future, it is important to consider how educational approaches and theories can and should adapt to both the capabilities of modern technology and the needs of the students within it. As more and more new technology, such as artificial intelligent (AI), and extended reality (XR) environments enter classrooms and educational setups – from primary school through to the university –, questions around their suitability arise. How can we promote better learning, or at the very least not impeded when using novel technology. This concern has been the main motivation behind the research conducted in this thesis. How to best incorporate new technology into learning environments is a challenge for researchers in the areas of HCI and education. This thesis, and the studies conducted within it, has provided a small contribution to meeting that challenge. Finally,

With or without "Magic"...
"Learning without reflection is a waste. Reflection without learning is dangerous."

-

Bibliography

[1] ©2022 IEEE. Reprinted, with permission, from M. Weerasinghe, A. Quigley, K. Č. Pucihar, A. Toniolo, A. Miguel and M. Kljun, Arigatō: Effects of Adaptive Guidance on Engagement and Performance in Augmented Reality Learning Environments, in IEEE Transactions on Visualization and Computer Graphics, 28(11): 3737-3747, 2022, doi: 10.1109/TVCG.2022.3203088.

[2] ©2022 IEEE. Reprinted, with permission, from M. Weerasinghe et al., VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations, in IEEE Transactions on Visualization and Computer Graphics, 28(11): 3748-3758, 2022, doi: 10.1109/TVCG.2022.3203116.

[3] W. K. Adams, A. Paulson, and C. E. Wieman. What levels of guidance promote engaged exploration with interactive simulations? *AIP conference proceedings*, 1064(1):59–62, 2008.

[4] S. Ahmed. A review on using opportunities of augmented reality and virtual reality in construction project management. *Organization, technology & management in construction: an international journal*, 10(1):1839–1852, 2018.

[5] W. S. Albert, R. A. Rensink, and J. M. Beusmans. Learning relative directions between landmarks in a desktop virtual environment. *Spatial cognition and computation*, 1(2):131–144, 1999.

[6] T. M. Alonzo, G. R. LaCagnina, and B. G. Olsen. Behaviorism vs. humanism: Two contrasting approaches to learning theory. *Southern Journal of Educational Research*, 1977.

[7] O. Alrashidi, H. P. Phan, and B. H. Ngu. Academic engagement: An overview of its definitions, dimensions, and major conceptualisations. *International Education Studies*, 9(12):41–52, 2016.

[8] E. A. Alrehaili and H. Al Osman. A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments*, 30(5):922–935, 2022.

[9] J. R. Anderson. *Rules of the mind*. Psychology Press, 2014.

[10] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent tutoring systems. *Science*, 228(4698):456–462, 1985.

[11] O. Anonthanasap, C. He, K. Takashima, T. Leelanupab, and Y. Kitamura. Mnemonic-based interactive interface for second-language vocabulary learning. *Proceedings of the Human Interface Society, HIS*, 14, 2014.

[12] J. J. Appleton, S. L. Christenson, and M. J. Furlong. Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5):369–386, 2008.

[13] F. Ashworth, G. Brennan, K. Egan, R. Hamilton, and O. Sáenz. Learning theories and higher education. *Level 3*, 2(1):4, 2004.

[14] R. C. Atkinson. Mnemotechnics in second-language learning. *American psychologist*, 30(8):821, 1975.

[15] R. T. Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.

[16] S. O. Bada and S. Olusegun. Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, 5(6):66–70, 2015.

[17] J. Birt, Z. Stromberga, M. Cowling, and C. Moro. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information*, 9(2):31, 2018.

[18] R. L. Blum and M. G. Walker. The lisp programming language and environments. *Implementing Health Care Information Systems*, pages 353–364, 1989.

[19] W. T. Botelho, M. d. G. B. Marietto, J. C. d. M. Ferreira, and E. P. Pimentel. Kolb's experiential learning theory and belhot's learning cycle guiding the use of computer simulation in engineering education: A pedagogical proposal to shift toward an experiential pedagogy. *Computer Applications in Engineering Education*, 24(1):79–88, 2016.

[20] D. Boud, R. Cohen, and D. Walker. *Using experience for learning*. McGraw-Hill Education (UK), 1993.

[21] D. Boud, R. Keogh, and D. Walker. *Reflection: Turning experience into learning*. Routledge, 2013.

[22] M. Bower, M. J. Lee, and B. Dalgarno. Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment. *British Journal of Educational Technology*, 48(2):407–430, 2017.

[23] D. A. Bowman and R. P. McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007.

[24] D. R. Brodbeck. *Spatial Learning*, pages 3166–3168. Springer US, Boston, MA, 2012.

[25] J. S. Bruner. The act of discovery. *Understanding Children*, pages 10–24, 1961.

[26] A. I. Canhoto and J. Murphy. Learning from simulation design to develop better experiential learning initiatives: An integrative approach. *Journal of Marketing Education*, 38(2):98–106, 2016.

[27] K. J. Carbonneau and S. C. Marley. Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *The Journal of Experimental Education*, 83(4):495–513, 2015.

[28] R. N. Carney, J. R. Levin, and M. E. Levin. Enhancing the psychology of memory by enhancing memory of psychology. *Teaching of Psychology*, 21(3):171–174, 1994.

[29] W. M. Carroll. Using worked examples as an instructional support in the algebra classroom. *Journal of educational psychology*, 86(3):360, 1994.

[30] S. R. Cavanagh. How to make your teaching more engaging. *The Chronicle of Higher Education*, 2019.

[31] Y.-C. Chen and P.-C. Chen. The effect of english popular songs on learning motivation and learning performance. *WHAMPOA-An Interdisciplinary Journal*, 56:13–28, 2009.

[32] R. E. Clark. How much and what type of guidance is optimal for learning from instruction. *Constructivist instruction: Success or failure*, pages 158–183, 2009.

[33] R. E. Clark. How much and what type of guidance is optimal for learning from instruction. *Constructivist instruction: Success or failure*, pages 158–183, 2009.

[34] H. Coates. A model of online and general campus-based student engagement. *Assessment & Evaluation in Higher Education*, 32(2):121–141, 2007.

[35] L. J. Cronbach and R. E. Snow. *Aptitudes and instructional methods: A handbook for research on interactions.* Irvington, 1977.

[36] D. P. Crowne and D. Marlowe. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 33:529–535, Sep 1960.

[37] C. S. C. Dalim, M. S. Sunar, A. Dey, and M. Billinghurst. Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, 134:44–64, 2020.

[38] G. J. Dean. Developing experiential learning activities for adult learners. *American Association for Adult and Continuing Education National Conference*, 1993.

[39] J. Dewey. 1997. experience and education. *Touchstone, New York*, 1938.

[40] J. Dunlosky, K. A. Rawson, E. J. Marsh, M. J. Nathan, and D. T. Willingham. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest*, 14(1):4–58, 2013.

[41] M. Eckert, J. S. Volmerg, C. M. Friedrich, et al. Augmented reality in medicine: systematic and bibliographic review. *JMIR mHealth and uHealth*, 7(4):e10967, 2019.

[42] N. C. Ellis. The psychology of foreign language vocabulary acquisition: Implications for call. *Computer Assisted Language Learning*, 8(2-3):103–128, 1995.

[43] I. Elola and A. Oskoz. Collaborative writing: Fostering foreign language and writing conventions development. 2010.

[44] G. Ferrero, F. Bichai, and M. Rusca. Experiential learning through role-playing: Enhancing stakeholder collaboration in water safety plans. *Water*, 10(2):227, 2018.

[45] J. Fredricks, W. McColskey, J. Meli, J. Mordica, B. Montrosse, and K. Mooney. Measuring student engagement in upper elementary through high school: A description of 21 instruments. issues & answers. rel 2011-no. 098. *Regional Educational Laboratory Southeast*, 2011.

[46] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109, 2004.

[47] R. Fry and D. Kolb. Experiential learning theory and learning experiences in liberal arts education. *New directions for experiential learning*, 6:79, 1979.

[48] A. C. Graesser, M. W. Conley, and A. Olney. Intelligent tutoring systems. *APA educational psychology handbook, Vol 3: Application to learning and teaching.*, pages 451–473, 2012.

[49] J.-Q. Guan, L.-H. Wang, Q. Chen, K. Jin, and G.-J. Hwang. Effects of a virtual reality-based pottery making approach on junior high school students' creativity and learning engagement. *Interactive Learning Environments*, pages 1–17, 2021.

[50] J. W. Hall. On the utility of the keyword mnemonic for vocabulary learning. *Journal of Educational Psychology*, 80(4):554, 1988.

[51] B. Herbert, A. Weerasinghe, B. Ens, and M. Billinghurst. An adaptive ar tutor for cabling a network topology. *ICAT-EGVE*, pages 11–12, 2017.

[52] P. Herman and L. M. Gomez. Taking guided learning theory to school: Reconciling the cognitive, motivational, and social contexts of instruction. pages 74–93, 2009.

[53] A. Hinderks, M. Schrepp, and J. Thomaschewski. A benchmark for the short version of the user experience questionnaire. In *WEBIST*, pages 373–377, 2018.

[54] N. Horvat, S. Škec, T. Martinec, F. Lukačević, and M. M. Perišić. Comparing virtual reality and desktop interface for reviewing 3d cad models. In *Proceedings of the Design Society: International Conference on Engineering Design*, volume 1, pages 1923–1932. Cambridge University Press, 2019.

[55] M.-C. Hsieh and Y.-H. Lin. Vr and ar applications in medical practice and education. *Hu Li Za Zhi*, 64(6):12–18, 2017.

[56] T.-C. Hsu. Learning english with augmented reality: Do learning styles matter? *Computers & Education*, 106:137–149, 2017.

[57] T.-C. Huang. Seeing creativity in an augmented experiential learning environment. *Universal Access in the Information Society*, pages 1–13, 2019.

[58] T.-C. Huang, C.-C. Chen, and Y.-W. Chou. Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education*, 96:72–82, 2016.

[59] P. Jansen-Osmann and P. Fuchs. Wayfinding behavior and spatial knowledge of adults and children in a virtual environment: The role of landmarks. *Experimental Psychology*, 53(3):171–181, 2006.

[60] L. Jarmon, T. Traphagan, M. Mayrath, and A. Trivedi. Virtual world teaching, experiential learning, and assessment: An interdisciplinary communication course in second life. *Computers & Education*, 53(1):169–182, 2009.

[61] E. Jensen. *Brain-based learning: The new paradigm of teaching*. Corwin Press, 2008.

[62] L. Joplin. On defining experiential education. *Journal of experiential education*, 4(1):17–20, 1981.

[63] S. Kalyuga, P. Chandler, and J. Sweller. Learner experience and efficiency of instructional guidance. *Educational Psychology*, 21(1):5–23, 2001.

[64] D. Kamińska, T. Sapiński, S. Wiak, T. Tikk, R. E. Haamer, E. Avots, A. Helmi, C. Ozcinar, and G. Anbarjafari. Virtual reality and its applications in education: Survey. *Information*, 10(10):318, 2019.

[65] M. T. Keeton and P. J. Tate. *Learning by experience–what, why, how*. Number 1. Jossey-Bass, 1978.

[66] M. E. King-Sears, C. D. Mercer, and P. T. Sindelar. Toward independence with keyword mnemonics: A strategy for science vocabulary instruction. *Remedial and Special Education*, 13(5):22–33, 1992.

[67] P. A. Kirschner, J. Sweller, and R. E. Clark. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2):75–86, 2006.

[68] K. R. Koedinger and V. Aleven. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3):239–264, 2007.

[69] D. Kolb. *Experiential learning: experience as the source of learning and development*. Prentice Hall, 1984.

[70] D. A. Kolb. Experiential learning theory and the learning style inventory: A reply to freedman and stumpf. *Academy of Management Review*, 6(2):289–296, 1981.

[71] D. A. Kolb, R. E. Boyatzis, C. Mainemelis, et al. Experiential learning theory: Previous research and new directions. *Perspectives on thinking, learning, and cognitive styles*, 1(8):227–247, 2001.

[72] R. Kop and A. Hill. Connectivism: Learning theory of the future or vestige of the past? *International Review of Research in Open and Distributed Learning*, 9(3):1–13, 2008.

[73] M. Kozhevnikov, M. A. Motes, and M. Hegarty. Spatial visualization in physics problem solving. *Cognitive science*, 31(4):549–579, 2007.

[74] G. D. Kuh. What student affairs professionals need to know about student engagement. *Journal of college student development*, 50(6):683–706, 2009.

[75] D. Kuhn. Is direct instruction an answer to the right question? *Educational psychologist*, 42(2):109–113, 2007.

[76] C. Kwon. Verification of the possibility and effectiveness of experiential learning using hmd-based immersive vr technologies. *Virtual Reality*, 23(1):101–118, 2019.

[77] F. Lateef. Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma and Shock*, 3(4):348, 2010.

[78] A. W. Lazonder. Inquiry learning. In *Handbook of research on educational communications and technology*, pages 453–464. Springer, 2014.

[79] A. W. Lazonder and R. Harmsen. Meta-analysis of inquiry-based learning: Effects of guidance. *Review of educational research*, 86(3):681–718, 2016.

[80] A. Lele. Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4:17–26, 2013.

[81] L. H. Lewis and C. J. Williams. Experiential learning: Past and present. *New directions for adult and continuing education*, 1994(62):5–16, 1994.

[82] S. Li, Y. Chen, D. M. Whittinghill, and M. Vorvoreanu. A pilot study exploring augmented reality to increase motivation of chinese college students learning english. In *2014 ASEE Annual Conference & Exposition*, pages 24–85, 2014.

[83] M.-H. Lin, H.-g. Chen, et al. A study of the effects of digital learning on learning motivation and learning outcome. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7):3553–3564, 2017.

[84] M. A. Mastropieri and T. E. Scruggs. *Teaching students ways to remember: Strategies for learning mnemonically.* Brookline Books, 1991.

[85] S. Mathan and K. R. Koedinger. Recasting the feedback debate: Benefits of tutoring error detection and correction. *Artificial intelligence in education: Shaping the future of learning through intelligent technologies*, 922(6389):13, 2003.

[86] R. E. Mayer. Constructivism as a theory of learning versus constructivism as a prescription for instruction. *Constructivist instruction: Success or failure*, pages 184–200, 2009.

[87] E. Menaker, S. Coleman, J. Collins, and M. Murawski. Harnessing experiential learning theory to achieve warfighting excellence. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 2006.

[88] B. Mergel. Instructional design and learning theory, 1998.

[89] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[90] C. S. Miller, J. F. Lehman, and K. R. Koedinger. Goals and learning in microworlds. *Cognitive Science*, 23(3):305–336, 1999.

[91] S. Minocha and A. J. Reeves. Design of learning spaces in 3d virtual worlds: an empirical investigation of second life. In *Learning and Research in Virtual Worlds*, pages 19–45. Routledge, 2013.

[92] J. Moore. Behaviorism. *The Psychological Record*, 61:449–463, 2011.

[93] D. Mueller and J. M. Ferreira. Marvel: A mixed-reality learning environment for vocational training in mechatronics. In *Proceedings of the Technology Enhanced Learning International Conference (TEL 03)*, 2003.

[94] R. J. Nadolski, P. A. Kirschner, and J. J. Van Merriënboer. Optimizing the number of steps in learning tasks for complex skills. *British Journal of Educational Psychology*, 75(2):223–237, 2005.

[95] A. Nisansala, M. Weerasinghe, G. Dias, D. Sandaruwan, C. Keppitiyagama, N. Kodikara, C. Perera, and P. Samarasinghe. Flight simulator for serious gaming. In *Information Science and Applications*, pages 267–277. Springer, 2015.

[96] 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.

[97] J. Parong and R. E. Mayer. Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, 2020.

[98] W. Pfeiffer and J. Jones. A handbook of structured experiences for human relations. *San Diego: University Associates*, 1, 1983.

[99] M. Pressley, J. R. Levin, and H. D. Delaney. The mnemonic keyword method. *Review of Educational Research*, 52(1):61–91, 1982.

[100] M. Pressley, J. R. Levin, N. A. Kuiper, S. L. Bryant, and S. Michener. Mnemonic versus nonmnemonic vocabulary-learning strategies: Additional comparisons. *Journal of Educational Psychology*, 74(5):693, 1982.

[101] A. L. Putnam. Mnemonics in education: Current research and applications. *Translational Issues in Psychological Science*, 1(2):130, 2015.

[102] M. R. Raugh and R. C. Atkinson. A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 67(1):1, 1975.

[103] M. R. Raugh, R. D. Schupbach, and R. C. Atkinson. Teaching a large russian language vocabulary by the mnemonic keyword method. *Instructional science*, 6(3):199–221, 1977.

[104] J. Reeve and C.-M. Tseng. Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4):257–267, 2011.

[105] J. T. Richardson and A. Jones. *Mental imagery and human memory.* Macmillan International Higher Education, 1980.

[106] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56, 2007.

[107] N. Sagarra and M. Alba. The key is in the keyword: L2 vocabulary learning methods with beginning learners of spanish. *The modern language journal*, 90(2):228–243, 2006.

[108] SAGE Publications Sage CA: Los Angeles, CA. *Engagement and Competence in VR and non-VR Environments*, volume 61, 2017.

[109] G. Salomon. The differential investment of mental effort in learning from different sources. *Educational psychologist*, 18(1):42–50, 1983.

[110] G. Salomon. Television is" easy" and print is" tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of educational psychology*, 76(4):647, 1984.

[111] M. E. C. Santos, T. Taketomi, G. Yamamoto, M. M. T. Rodrigo, C. Sandor, H. Kato, et al. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1):1–23, 2016.

[112] W. B. Schaufeli. What is engagement. *Employee engagement in theory and practice*, 15:321, 2013.

[113] F. C. S. Schiller. *Humanism; philosophical essays*. Macmillan and Company, limited, 1912.

[114] H. G. Schmidt. Problem-based learning: Rationale and description. *Medical education*, 17(1):11–16, 1983.

[115] D. L. Schwartz and J. D. Bransford. A time for telling. *Cognition and instruction*, 16(4):475–5223, 1998.

[116] A. M. Shapiro and D. L. Waters. An investigation of the cognitive processes underlying the keyword method of foreign vocabulary learning. *Language teaching research*, 9(2):129–146, 2005.

[117] M. J. Soler and J. C. Ruiz. The spontaneous use of memory aids at different educational levels. *Applied Cognitive Psychology*, 10(1):41–51, 1996.

[118] L. P. Steffe and J. E. Gale. *Constructivism in education*. Psychology Press, 1995.

[119] C.-H. Su and T.-W. Cheng. A sustainability innovation experiential learning model for virtual reality chemistry laboratory: An empirical study with pls-sem and ipma. *Sustainability*, 11(4):1027, 2019.

[120] J. Sweller. Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier, 2011.

[121] S. Ternier, R. Klemke, M. Kalz, P. Van Ulzen, and M. Specht. Arlearn: Augmented reality meets augmented virtuality. *J. UCS*, 18(15):2143–2164, 2012.

[122] W. Thabet, M. F. Shiratuddin, and D. Bowman. Virtual reality in construction: a review. *Engineering computational technology*, pages 25–52, 2002.

[123] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators & Virtual Environments*, 13(5):560–571, 2004.

[124] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-

directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 161–174, 2019.

[125] S. Tobias. When do instructional methods. *Educational Researcher*, 11(4):4–9, 1982.

[126] S. Tobias and T. M. Duffy. *Constructivist instruction: Success or failure?* Routledge, 2009.

[127] E. C. Tolman. A new formula for behaviorism. *Psychological Review*, 29(1):44, 1922.

[128] F. B. Topu and Y. Goktas. The effects of guided-unguided learning in 3d virtual environment on students' engagement and achievement. *Computers in Human Behavior*, 92:1–10, 2019.

[129] J. G. Trafton and B. J. Reiser. The contributions of studying examples and solving problems to skill acquisition, 1994.

[130] J. G. Van Hell and A. C. Mahn. Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language learning*, 47(3):507–546, 1997.

[131] K. VanLehn. Student modeling. *Foundations of intelligent tutoring systems*, 55:78, 1988.

[132] C. D. Vazquez, A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2172–2179, 2017.

[133] S. Ventura, R. M. Baños, C. Botella, and N. Mohamudally. Virtual and augmented reality: New frontiers for clinical psychology. *State of the art virtual reality and augmented reality knowhow*, 10, 2018.

[134] D. Waller, A. C. Beall, and J. M. Loomis. Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology*, 24(1):105–116, 2004.

[135] M. Weerasinghe, K. Čopič Pucihar, J. Ducasse, A. Quigley, A. Toniolo, A. Miguel, N. Caluya, and M. Kljun. Exploring the future building: representational effects on projecting oneself into the future office space. *Virtual Reality*, 27(1):51–70, 2023.

[136] Z. Wei. Does teaching mnemonics for vocabulary learning make a difference? putting the keyword method and the word part technique to the test. *Language Teaching Research*, 19(1):43–69, 2015.

[137] G. Westerfield, A. Mitrovic, and M. Billinghurst. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, 25(1):157–172, 2015.

[138] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015.

[139] A. F. Wise and K. O'Neill. Beyond more versus less: A reframing of the debate on instructional guidance. 2009.

[140] B. G. Witmer and W. J. Sadowski Jr. Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human factors*, 40(3):478–488, 1998.

[141] J. B. Worthen and R. R. Hunt. *Mnemonology: Mnemonics for the 21st century*. Psychology Press, 2011.

[142] K. Yilmaz. The cognitive perspective on learning: Its theoretical underpinnings and implications for classroom practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 84(5):204–212, 2011.

[143] J. Zhao, P. LaFemina, J. Carr, P. Sajjadi, J. O. Wallgrün, and A. Klippel. Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based stem education. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 893–902. IEEE, 2020.

Appendix A

Study I: VRNav

A.1 Ethical Approvals

University of Primorska

Univerza na Primorskem
Università del Litorale
University of Primorska
Titov trg 4 • SI-6000 Koper
+386 (0)5 611 75 53
info@upr.si • www.upr.si

Koper, 19 August 2019

ETHICAL APPROVAL FOR THE RESEARCH TITLED:

"The effects of exploring the future building in VR, desktop and paper"

UP employees and students are obliged to conduct research in accordance with the UP Code of Ethics.

For the field of scientific research, the university is following the *European Code of Conduct for Research Integrity* issued by ALLEA, the *Association of European Academies*, in 2017. It is a widely recognised framework in the EU for self-regulation in all areas of research and is recognised by the European Commission as a reference document for research integrity in all EU-funded projects and as a model for organisations and researchers across Europe.

The aforementioned research titled "Exploring the future building: Representational effects on projecting oneself into the future office space" presents a minimal risk to users, who will experience three different representations of the same building and solve questionnaires. Standard university's consent form will be used as well as standard procedures to data anonymisation and storage. The researchers set the procedures in place to explain the aim of the study, its process, and to mitigate any possible dificullties. The researchers adhered to the accepted ethical standards of a genuine research study.

Rector,

Prof. dr. Klavdija Kutnar

KLAVDIJA KUTNAR
Digitally signed by KLAVDIJA KUTNAR
Date: 2022.07.16 15:40:50 +02'00'

University of St.Andrews

University of St Andrews | FOUNDED 1413

School of Computer Science Ethics Committee

27 July 2022

Dear Anuradhi,

Thank you for submitting your ethical application which was considered at the School Ethics Committee.

The School of Computer Science Ethics Committee, acting on behalf of the University Teaching and Research Ethics Committee (UTREC), has approved this application:

Approval Code:	CS16443	**Approved on:**	27.07.22	**Approval Expiry:**	27.07.27
Project Title:	The Effects of Exploring the Future Building in VR, Desktop and Paper				
Researcher(s):	Anuradhi Maheshya Weerasinghe				
Supervisor(s):	Dr Alice Toniolo and Dr Angela Miguel				

The following supporting documents are also acknowledged and approved:

1. Application Form
2. Participant Information and Consent Form
3. Questionnaires
4. FAMNIT Approval

Approval is awarded for 5 years, see the approval expiry data above.

If your project has not commenced within 2 years of approval, you must submit a new and updated ethical application to your School Ethics Committee.

If you are unable to complete your research by the approval expiry date you must request an extension to the approval period. You can write to your School Ethics Committee who may grant a discretionary extension of up to 6 months. For longer extensions, or for any other changes, you must submit an ethical amendment application.

You must report any serious adverse events, or significant changes not covered by this approval, related to this study immediately to the School Ethics Committee.

Approval is given on the following conditions:

- that you conduct your research in line with:
  - the details provided in your ethical application
  - the University's Principles of Good Research Conduct
  - the conditions of any funding associated with your work
- that you obtain all applicable additional documents (see the 'additional documents' webpage for guidance) before research commences.

You should retain this approval letter with your study paperwork.

School of Computer Science Ethics Committee
Dr Juan Ye/Convenor, Jack Cole Building, North Haugh, St Andrews, Fife, KY16 9SX
Telephone: 01334 463252 Email: ethics-cs@st-andrews.ac.uk

A.2 Participant Consent Form

Informed consent form for the study about immersiveness in VR

The study is organised by the:

- **HICUP Lab - Laboratory for studying Humans Interacting with Computers at University of Primorska**
- **InnoRenew CoE, a not-profit private research institute** and the
- **SACHI, the St Andrews Computer Human Interaction research group**

Certificate of consent

Please read the ten statements below and tick the boxes to confirm your agreement.	☑
I have read all sections in this information sheet.	☐
I have been given the opportunity to ask questions about the project.	☐
I agree to take part in the study which includes audio and video recording, eye-tracking.	☐
I agree to take part in the study which collects eye tracking, electrodermal and heart rate activity.	☐
I understand that interaction with interface will be recorded.	☐
I understand my participation is voluntary, and I can withdraw at any time.	☐
I understand my words may be quoted in publications and other research outputs.	☐
I understand my real name will be removed in all publications and outputs.	☐
I understand my personal data will be kept securely and available only to authorized personnel.	☐
I understand anonymised research data will be archived and may be used by third parties.	☐
I assign the copyright I hold to the content generated in this activity to the organisers.	☐

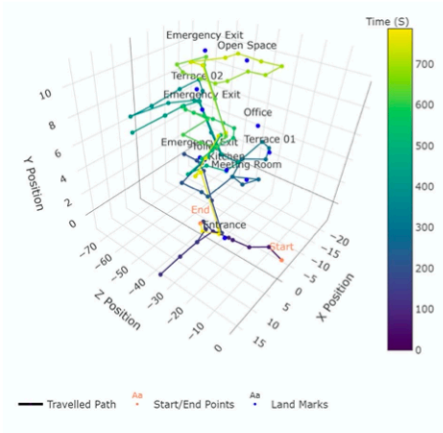
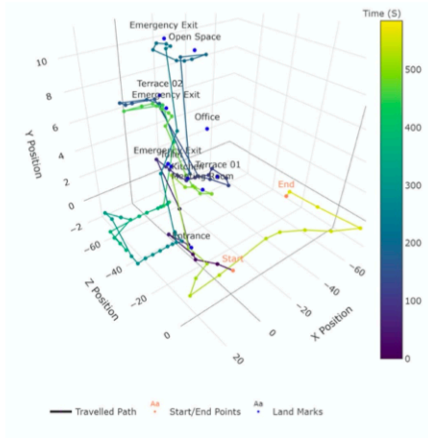
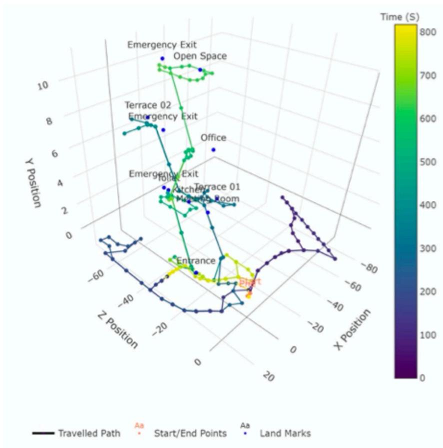
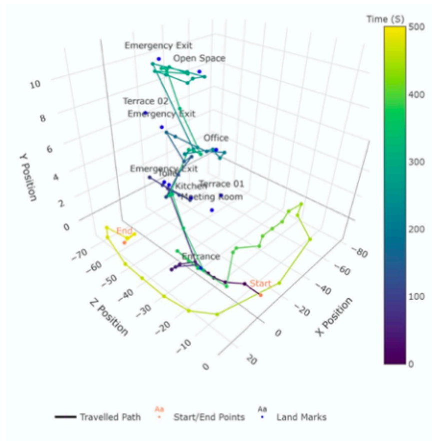
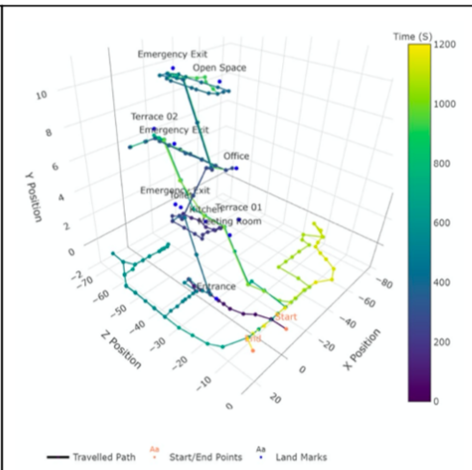
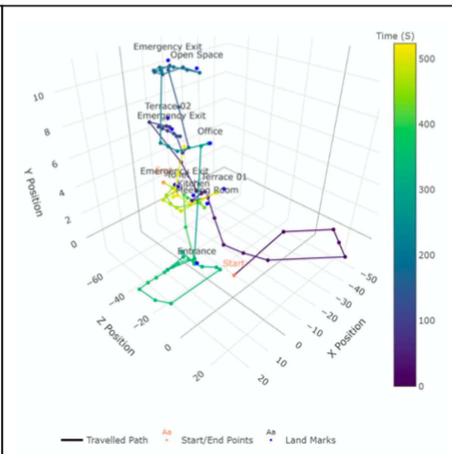
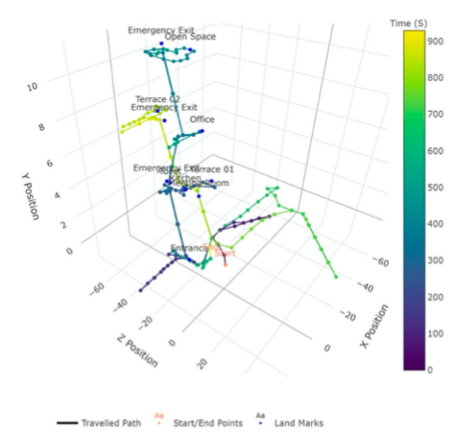
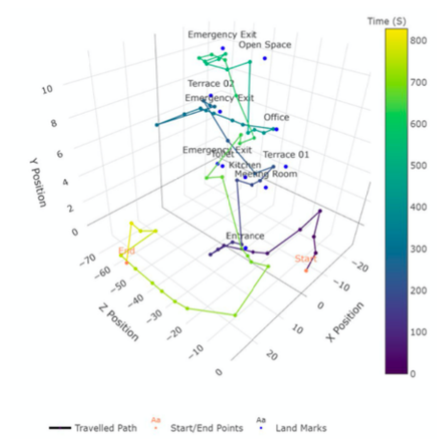
I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have asked were answered to my satisfaction. I consent voluntarily to be a participant in this study.

Print name of participant	Signature of participant	Date

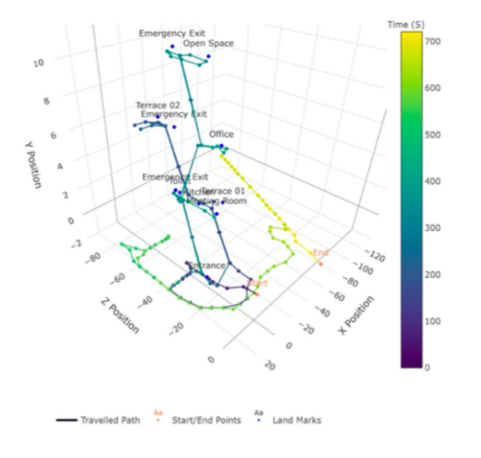
I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered appropriately and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and that consent has been given freely and voluntarily. A copy of this certificate of consent has been provided to the participant.

Print name of researcher taking the consent	Signature of participant	Date

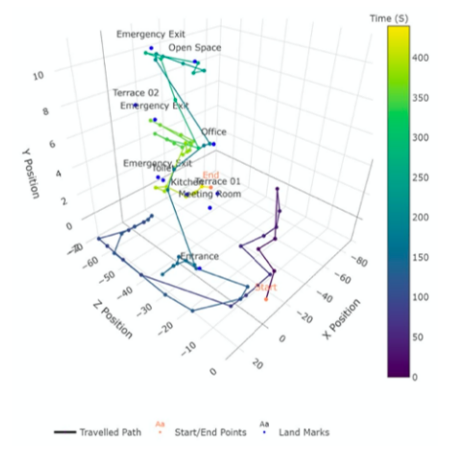
A.3 Additional Results

<table>
<tr><th>User</th><th>VR</th><th>Desktop</th></tr>
<tr><td>01</td><td>
</td><td>
</td></tr>
<tr><td>02</td><td>
</td><td>
</td></tr>
<tr><td>03</td><td>
</td><td>
</td></tr>
<tr><td>04</td><td>
</td><td>
</td></tr>
</table>

05

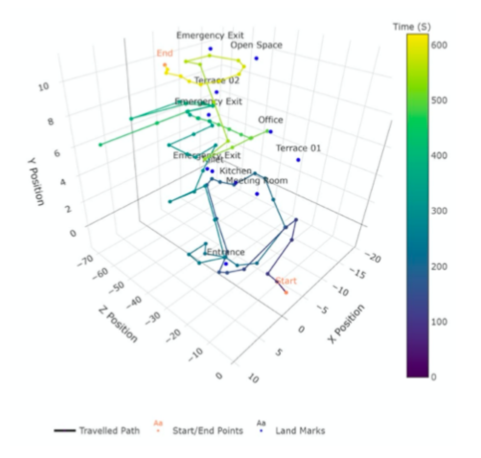


Time (S)
Emergency Exit
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Kitchen
Terrace 01
Meeting Room
Entrance
Start
End
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

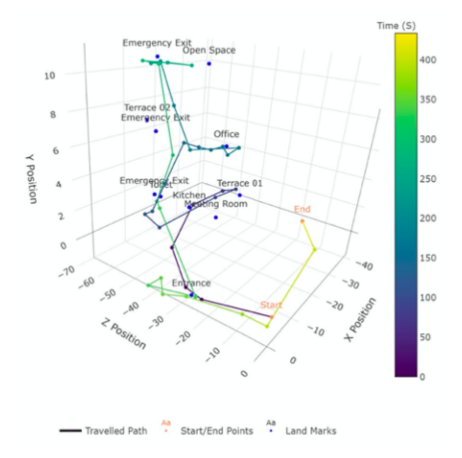


Time (S)
Emergency Exit
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Kitchen
Terrace 01
Meeting Room
Entrance
Start
End
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

06

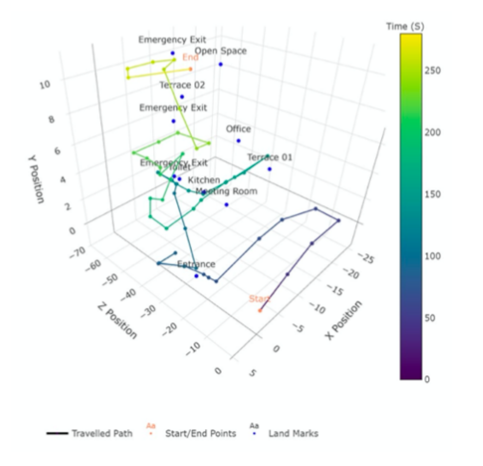


Time (S)
Emergency Exit
Open Space
End
Terrace 02
Emergency Exit
Office
Terrace 01
Emergency Exit
Kitchen
Meeting Room
Entrance
Start
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

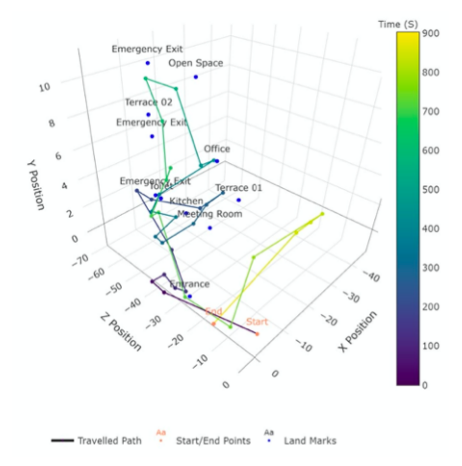


Time (S)
Emergency Exit
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Terrace 01
Kitchen
Meeting Room
End
Entrance
Start
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

07

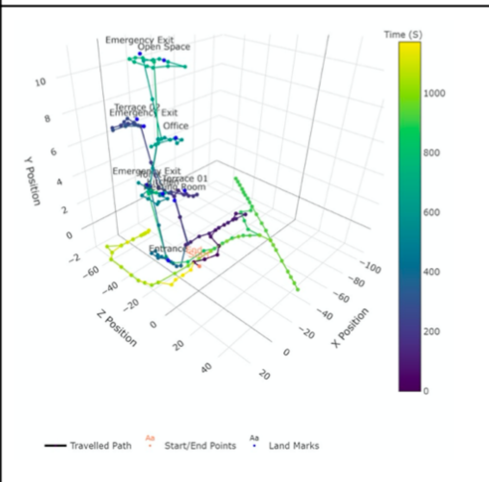


Time (S)
Emergency Exit
End
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Terrace 01
Kitchen
Meeting Room
Entrance
Start
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

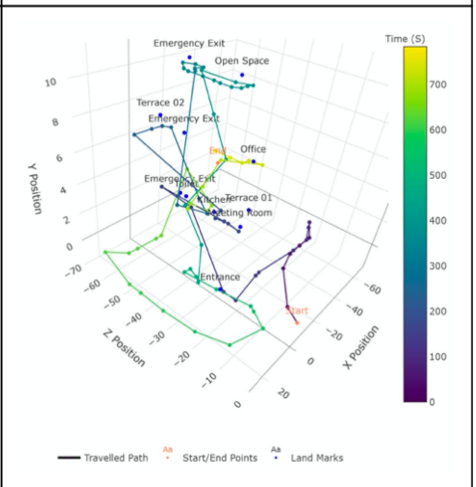


Time (S)
Emergency Exit
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Terrace 01
Kitchen
Meeting Room
Entrance
End
Start
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

08



Time (S)
Emergency Exit
Open Space
Emergency Exit
Office
Emergency Exit
Terrace 01
Entrance
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks



Time (S)
Emergency Exit
Open Space
Terrace 02
Emergency Exit
Office
Emergency Exit
Kitchen
Terrace 01
Meeting Room
Entrance
Start
Y Position
Z Position
X Position
Travelled Path
Start/End Points
Land Marks

Appendix B

Study II: Arigatō

B.1 Ethical Approvals

University of Primorska

Research ethics committee
Department of Psychology UP FAMNIT
Glagoljaška ulica 8
6000 Koper

Date: 2 September 2021

Application number: **2021-12**

Maheshya Weerasinghe,
UP FAMNIT, Department of Information Sciences and Technologies
maheshya@famnit.upr.si

SUBJECT: Expert opinion of the Committee on the study proposal

Dear Ms Weerasinghe,

The Research ethics committee of the Department of Psychology, UP FAMNIT, has received and considered your application for the review of the study proposal entitled *"Vloga navodil pri učenju v razširjeni resničnosti / Instructional Guidance in Extended Reality for Learning"*.

On 26 August 2021, the Committee decided that the application for the review of the study proposal was adequately prepared and concluded that **the study proposal complies with the research ethics principles**.

With kind regards,

dr. Urša Mars Bitenc,
Chair of the Research ethics committee
Department of Psychology UP FAMNIT

University of St.Andrews

University of St Andrews | FOUNDED 1413

School of Computer Science Ethics Committee

06 October 2021

Dear Anuradhi,

Thank you for submitting your ethical application which was considered at the School Ethics Committee on Wednesday 6th October.

The School of Computer Science Ethics Committee, acting on behalf of the University Teaching and Research Ethics Committee (UTREC), has approved this application:

Approval Code:	CS15762	**Approved on:**	06.10.2021	**Approval Expiry:**	06.10.2026
Project Title:	Instructional Guidance in Extended Reality for Learning				
Researcher(s):	Anuradhi Maheshya Weerasinghe				
Supervisor(s):	Alice Toniolo an Angela Miguel				

The following supporting documents are also acknowledged and approved:

1. Application Form
2. Participant Information Sheet
3. Participant Consent Form
4. Participant Debrief Form
5. Advert
6. Questionnares
7. Approval Document from FAMNIT

Approval is awarded for 5 years, see the approval expiry data above.

If your project has not commenced within 2 years of approval, you must submit a new and updated ethical application to your School Ethics Committee.

If you are unable to complete your research by the approval expiry date you must request an extension to the approval period. You can write to your School Ethics Committee who may grant a discretionary extension of up to 6 months. For longer extensions, or for any other changes, you must submit an ethical amendment application.

You must report any serious adverse events, or significant changes not covered by this approval, related to this study immediately to the School Ethics Committee.

Approval is given on the following conditions:

- that you conduct your research in line with:
  - the details provided in your ethical application
  - the University's Principles of Good Research Conduct
  - the conditions of any funding associated with your work
- that you obtain all applicable additional documents (see the 'additional documents' webpage for guidance) before research commences.

You should retain this approval letter with your study paperwork.

School of Computer Science Ethics Committee
Dr Juan Ye/Convenor, Jack Cole Building, North Haugh, St Andrews, Fife, KY16 9SX
Telephone: 01334 463252 Email: ethics-cs@st-andrews.ac.uk

B.2 Participant Consent Form

Consent Form

Instructional Guidance in Extended Reality for Learning

Maheshya Weerasinghe, Matjaž Kljun, Alice Toniolo, Klen Čopič Pucihar, Angela Miguel, Aaron Quigley

We kindly ask you to consider the following points before signing this form. Your signature confirms that you are willing to participate in this study, however, signing this form does not commit you to anything you do not wish to do and you are free to withdraw your participation at any time.

Please initial box

- I understand the contents of the Participant Information Sheet ☐
- I have been given the opportunity to ask questions about the study and have had them answered satisfactorily. ☐
- I understand that my participation is entirely voluntary and that I can withdraw from the study at any time without giving an explanation and with no disbenefit. ☐
- I understand the precautions that will be in place to reduce the risk of coronavirus and how I can help reduce this risk. ☐
- I understand who will have access to my data, how it will be stored, in what form it will be shared, and what will happen to it at the end of the study. ☐
- I understand my words may be quoted in publications and other research outputs. ☐
- I understand my real name will be removed in all publications and outputs. ☐
- I understand that I will be able to withdraw my data at any time, and I understand that if my data has been anonymised, it cannot be withdrawn. ☐
- I assign the copyright I hold to the content generated in this activity to the organizers. ☐
- I agree to take part in the above study ☐

GSR /HMD screen recording

I understand that part of this research involves recording GSR sensor data/HMD screen data.

- I agree to take part in the study which collects electrodermal and heart rate activity. ☐
- I understand that the HMD Screen and the interactions with the interface will be recorded. ☐

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have asked were answered to my satisfaction. I consent voluntarily to be a participant in this study.

Print name of participant	Signature of participant	Date

Appendix C

Study III: Vocabulary

C.1 Ethical Approvals

University of Primorska

Komisija Univerze na Primorskem za etiko v raziskavah, ki vključujejo delo z ljudmi (KER UP)
Commission of the University of Primorska for Ethics in Human Subjects Research (KER UP)

UNIVERZA NA PRIMORSKEM
UNIVERSITÀ DEL LITORALE
UNIVERSITY OF PRIMORSKA
Titov trg 4, SI – 6000 Koper
Tel.: + 386 5 611 75 00
Fax.: + 386 5 611 75 30
E-mail: info@upr.si
http://www.upr.si

izr. prof. dr. Matjaž Kljun / Assoc. Prof. Matjaž Kljun
UP FAMNIT

Naša št. /N°: 4264-10-3/22
Koper, 31. 3. 2022 /31st March 2022

ETIČNO SOGLASJE K RAZISKAVI »Učenje besednega zaklada na podlagi asociacij (z besedo in besedo ter objektom) z in brez dopolnjene resničnosti«

Spoštovani.

Komisija Univerze na Primorskem za etiko v raziskavah, ki vključujejo delo z ljudmi (KER UP) je dne 9. 2. 2022 sprejela v obravnavo vašo vlogo za etično presojo raziskave »Učenje besednega zaklada na podlagi asociacij (z besedo in besedo ter objektom) z in brez dopolnjene resničnosti«.

KER UP je na 12. redni seji dne 29. 3. 2022 obravnavala vašo vlogo ter presodila, da je raziskava etično sprejemljiva. S tem vam za njeno izvedbo izdaja soglasje.

V raziskavi ste dolžni uporabljati tekst iz potrjenega obrazca za soglasje udeleženca za sodelovanje v raziskavi, ki vam ga posredujemo v priponki. Pri uporabi instrumentov in postopkov pa je potrebno posebno skrb nameniti upoštevanju načela minimalnega zbiranja osebnih podatkov ter zagotavljanju anonimnosti podatkov.

Lep pozdrav,

ETHICAL APPROVAL FOR THE RESEARCH »Vocabulary learning based on associations (with keyword and visualisation) with and without augmented reality«

Dear Madam / Sir,

On 9th February 2022, the Commission of the University of Primorska for Ethics in Human Subjects Research (KER UP) accepted your application for ethical review of the research entitled »Vocabulary learning based on associations (with keyword and visualisation) with and without augmented reality«.

KER UP assessed your application on its 12th regular meeting on 29th March 2022 and decided that the research is ethically acceptable. Therefore KER UP grants consent to the implementation of the research.

Within the research, you are obliged to use text from the approved form for the informed consent of the participant to participate in the research. The form is attached to this letter. When applying instruments and procedures, special care should be taken in order to follow personal data minimization principle and ensure the data anonymization.

Yours sincerely,

Pripravila / Prepared by:

Kristina Maršič
Skrbnica KER UP
Administator of KER UP

VITA POŠTUVAN Digitally signed by VITA POŠTUVAN Date: 2022.03.31 11:45:03 +02'00'

izr. prof. dr. Vita Poštuvan
Predsednica KER UP
President of KER UP

University of St.Andrews

School of Computer Science Ethics Committee

21 April 2022

Dear Anuradhi,

Thank you for submitting your ethical application which was considered by the School Ethics Committee.

The School of Computer Science Ethics Committee, acting on behalf of the University Teaching and Research Ethics Committee (UTREC), has approved this application:

Approval Code:	CS16151	**Approved on:**	21.04.2022	**Approval Expiry:**	21.04.2027
Project Title:	Vocabulary learning based on associations with and without augmented reality				
Researcher(s):	Anuradhi Maheshya Weerasinghe				
Supervisor(s):	Dr Alice Toniolo and Dr Angela Miguel				

The following supporting documents are also acknowledged and approved:

1. Application Form
2. Participant Information Sheet
3. Participant Consent Form
4. Participant Debrief Form
5. Advertisement
6. Questionnaire(s)
7. External Approval

Approval is awarded for 5 years, see the approval expiry data above.

If your project has not commenced within 2 years of approval, you must submit a new and updated ethical application to your School Ethics Committee.

If you are unable to complete your research by the approval expiry date you must request an extension to the approval period. You can write to your School Ethics Committee who may grant a discretionary extension of up to 6 months. For longer extensions, or for any other changes, you must submit an ethical amendment application.

You must report any serious adverse events, or significant changes not covered by this approval, related to this study immediately to the School Ethics Committee.

Approval is given on the following conditions:

- that you conduct your research in line with:
  - the details provided in your ethical application
  - the University's Principles of Good Research Conduct
  - the conditions of any funding associated with your work
- that you obtain all applicable additional documents (see the 'additional documents' webpage for guidance) before research commences.

You should retain this approval letter with your study paperwork.

School of Computer Science Ethics Committee
Dr Juan Ye/Convenor, Jack Cole Building, North Haugh, St Andrews, Fife, KY16 9SX
Telephone: 01334 463252 Email: ethics-cs@st-andrews.ac.uk

C.2 Participant Consent Form

Consent Form

VocabulARy: Learning Vocabulary in AR Supported by Keyword Visualisations

Maheshya Weerasinghe, Alice Toniolo, Angela Miguel, Matjaž Kljun, Klen Čopič Pucihar, Aaron Quigley

We kindly ask you to consider the following points before signing this form. Your signature confirms that you are willing to participate in this study, however, signing this form does not commit you to anything you do not wish to do, and you are free to withdraw your participation at any time.

Please initial box

- I understand the contents of the Participant Information Sheet ☐
- I have been given the opportunity to ask questions about the study and have had them answered satisfactorily. ☐
- I understand that my participation is entirely voluntary and that I can withdraw from the study at any time without giving an explanation and with no disbenefit. ☐
- I understand the precautions that will be in place to reduce the risk of coronavirus and how I can help reduce this risk. ☐
- I understand who will have access to my data, how it will be stored, in what form it will be shared, and what will happen to it at the end of the study. ☐
- I understand my words may be quoted in publications and other research outputs. ☐
- I understand my real name will be removed in all publications and outputs. ☐
- I understand that I will be able to withdraw my data at any time, and I understand that if my data has been anonymised, it cannot be withdrawn. ☐
- I assign the copyright I hold to the content generated in this activity to the organizers. ☐
- I agree to take part in the above study ☐

GSR

I understand that part of this research involves recording GSR sensor data.

- I agree to take part in the study which collects electrodermal and heart rate activity. ☐

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have asked were answered to my satisfaction. I consent voluntarily to be a participant in this study.

Name of participant | Signature of participant | Date