SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Investigating Immersive Virtual Reality Skill Training with Focus on Training Performance, Haptic Feedback, Physiological Arousal, and Adaptive Training Strategies

PhD dissertation

Unnikrishnan Radhakrishnan

Aarhus BSS
Aarhus University
Department of Business Development and Technology
2023

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Konstantinos Koumaditis, PhD and Francesco Chinello, PhD, for their unwavering support, guidance, and invaluable insights throughout my PhD journey. Their dedication and knowledge have been invaluable, and I feel incredibly fortunate to have had the opportunity to learn from and work with such exceptional mentors.

My heartfelt appreciation goes to my colleagues from the Extended Reality and Robotics Lab, Lasse Lui and Pedro Gomez, as well as Lisheng Kuang and Claudio Pacchierotti from INRIA, for their camaraderie, collaboration and encouragement during my time at the lab and as a visiting researcher at INRIA. I sincerely thank the other members of the Department of Business Development and Technology, as well as the members of the Cognition and Behaviour Lab, for their support and for fostering a stimulating research environment.

On a personal note, I wish to express my profound love and gratitude to my father, T.M. Radhakrishnan, my uncle, Dr T.K. Gopalakrishnan Nair, my grandmother, Syamala Kumari, and my father-in-law, P.M. Chandran, who are no longer with us. Their memories and legacies have been a constant source of inspiration and motivation for me. I am immensely thankful to my guru, Sri Mata Amritanandamayi Devi, for providing spiritual guidance, inspiration, and strength throughout the course of my PhD. Her teachings and blessings have been instrumental in shaping my academic and personal growth.

I am forever grateful to all those who have contributed to my journey, directly or indirectly.

This work is dedicated to Appoppan, Amma, Jyotsna, Devi, Appu, and Sudha-amma for their unwavering love and support.

Table of Contents

Executive Summary

Immersive virtual reality (IVR) technologies allow users to feel fully engaged and absorbed within a virtual world using specialized hardware and software. The advent of these technologies has opened new possibilities for training and skill development across various industries. This dissertation explores the effectiveness of IVR for industrial skills training by conducting a comprehensive review of the academic literature and industry practices, as well as a series of controlled experiments to investigate methods for improving the effectiveness of skill training with IVR. This dissertation is guided by five main research questions:

RQ 1. What is the current state of the art in academic literature and industry practice regarding skills training using IVR?

RQ 2. Is IVR training effective compared to physical training?

RQ 3. What is the link between the physiological arousal level of the trainees and the effectiveness of IVR training?

RQ 4. Can haptic feedback make IVR training more effective?

RQ 5. What is the link between adaptive training and the effectiveness of IVR training?

The dissertation answers these five questions in the span of four chapters. The first chapter introduces the research background, motivation, and objectives, setting the stage for a detailed investigation into the role of IVR in skill training and its potential implications for industry applications. The second chapter addresses RQ1 by presenting a systematic literature review of academic research on industrial skills training and concludes with three case studies detailing the use of IVR for skill training in Denmark. This chapter is a collection of two articles[1][2] *"A Systematic Review of Immersive Virtual Reality for Industrial Skills Training"* and "*Immersive Virtual Reality Training: Three Cases from the Danish Industry."*

The third chapter addresses RQ 2 and RQ 3 by investigating the effectiveness of IVR for a fine motor skill training task using a buzz-wire game scenario, comparing it with physical training through a controlled experiment. Additionally, the experiment explores the link between physiological arousal, measured by skin conductance and heart rate variability, and improvements in performance that resulted from IVR training among the experiment subjects.

[1] Published in the journal Behaviour and Information Technology, Taylor & Francis.
[2] Published in the proceedings of the IEEE VR 2021 conference.

This chapter encompasses the article[3] titled *"Investigating the Effectiveness of Immersive VR Skill Training and its Link to Physiological Arousal."* The fourth chapter primarily examines RQ 4 with the help of an experiment the role of two haptic feedback modalities – kinesthetic and vibrotactile feedback – in improving the performance of subjects after undergoing IVR-based fine motor skill training in the buzz-wire task. This chapter also explores the link between physiological arousal and training performance (RQ 3) while using techniques adapted from chapter 3. This chapter contains the article[4] titled *"Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task."* The fifth chapter experimentally investigates the impact of adaptation based on subjects' self-efficacy on their performance in the buzz-wire scenario after undergoing IVR-based training.

Through empirical methods, this dissertation uncovers the approaches adopted by academia, and Danish companies, in addressing industrial skills training with IVR. This revealed trends in the use of IVR training for various procedural, decision-making, spatial, and fine/gross motor skills across industries and their potential to be adapted for remote training. Inspired by these findings, empirical studies in the form of IVR-based motor skill training experiments were designed and executed utilizing approximately three hundred subjects. The findings from these experiments highlight the possibilities, limitations, and challenges of using innovative technologies in IVR training scenarios involving the measurement of physiological arousal using biosensors, the implementation of haptic feedback, and adaptive training approaches. These experiments unambiguously indicated IVR training to be effective.

Overall, this dissertation contributes to the advancement of IVR for industrial skills training and offers guidance for future development and implementation, paving the way for designing more effective IVR-based skill training solutions applicable across a wide range of industries.

[3] Published in the journal "Virtual Reality", Springer.
[4] Under review

Resumé

Fremkomsten af teknologier inden for *immersive virtual reality* (IVR) har åbnet nye muligheder for træning og kompetenceudvikling på tværs af forskellige brancher. I denne afhandling undersøges effektiviteten af IVR i forbindelse med praktisk færdighedstræning. Undersøgelserne består af en omfattende gennemgang af akademisk litteratur på området og af hvad der sker i praksis samt en række kontrollerede eksperimenter. Formålet med dette er at undersøge metoder til at forbedre effektiviteten af færdighedstræning med IVR. Afhandlingens tema leder os igennem fem overordnede forskningsspørgsmål:

1. *Hvad er det nyeste tekniske niveau i den akademiske litteratur og i praksis vedrørende færdighedstræning ved hjælp af IVR?*
2. *Er IVR-træning effektiv sammenlignet med fysisk træning?*
3. *Hvad er sammenhængen mellem deltagernes fysiologiske ophidselsesniveau og IVR-træningseffektiviteten?*
4. *Hvad er sammenhængen mellem haptisk feedback-modalitet og IVR-træningseffektivitet?*
5. *Hvad er sammenhængen mellem adaptiv træning og IVR-effektivitet?*

Afhandlingens fire kapitler besvarer disse fem spørgsmål. I det første kapitel introduceres baggrunden, motivationen og formålene med forskningen, og her præsenteres også en detaljeret undersøgelse af IVR's rolle i færdighedstræning og dens muligheder for at blive anvendt i praksis. I det andet kapitel behandles det første forskningsspørgsmål via en systematisk litteraturgennemgang af akademisk forskning i praktisk færdighedstræning. Kapitlet afsluttes med tre casestudier, der beskriver, hvordan man har brugt VR til færdighedstræning i Danmark. Kapitlet består af to artikler *"A Systematic Review of Immersive Virtual Reality for Industrial Skills Training"* og *"Immersive Virtual Reality Training: Three Cases from the Danish Industry."*

Det tredje kapitel behandler forskningsspørgsmålene 2 og 3 ved i et kontrolleret eksperiment at undersøge effektiviteten af IVR til at løse en finmotorisk træningsopgave ved hjælp af et *buzz-wire*-spilscenarie og sammenligne det med fysisk/praktisk træning. I eksperimentet undersøges også sammenhængen mellem fysiologisk ophidselse, målt på huden og via pulsvariation, og forbedringer i forsøgspersonernes træningsresultater. Dette kapitel består af

artiklen *"Investigating the Effectiveness of Immersive VR Skill Training and its Link to Physiological Arousal."* I det fjerde kapitel undersøges ved hjælp af et eksperiment, hvilken rolle to haptiske feedback-modaliteter – kinæstetisk og vibrotaktil feedback – spiller med hensyn til at forbedre forsøgspersonernes præstationen efter at have gennemgået IVR-baseret finmotorisk træning i buzz-wire-opgaven. Dette kapitel indeholder artiklen *"Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task."* I det femte kapitel undersøges det eksperimentelt og via forsøgspersonernes tiltro til egne evner og performance, hvilken effekt tillæringen havde på deres præstationsforbedring i buzz-wire-scenariet efter at have gennemgået IVR-baseret træning.

Gennem empiriske metoder afdækker denne afhandling de tilgange, som den akademiske verden og danske virksomheder har brugt i deres tilgang til praktisk færdighedstræning med IVR. Dette afslørede tendenser i brugen af IVR-træning med hensyn til proceduremæssige, beslutningstagningsmæssige, rumlige og fin-/grovmotoriske færdigheder på tværs af brancher og tilgangenes potentiale til at kunne bruges til træning på distancen. Inspireret af disse resultater blev empiriske undersøgelser designet i form af IVR-baserede motoriske træningseksperimenter, og cirka tre hundrede forsøgspersoner deltog. Resultaterne af disse eksperimenter viser mulighederne, begrænsningerne og udfordringerne ved at bruge innovative teknologier i IVR-træningsscenarier med måling af fysiologisk ophidselse ved hjælp af biosensorer, implementering af haptisk feedback og adaptive træningstilgange. Disse eksperimenter indikerede utvetydigt, at IVR-træning var effektiv.

Samlet set bidrager denne afhandling til fremme af IVR til praktisk færdighedstræning, og resultaterne kan bruges til at skubbe den fremtidige udvikling og implementering i den rigtige retning, hvilket baner vejen for at designe mere effektive IVR-baserede løsninger til træning af færdigheder, der kan anvendes i en bred vifte af industrier.

List of figures

List of tables

Nomenclature

Abbreviations	
VR	Virtual Reality
IVR	Immersive VR
HMD	Head Mounted Display
GSR	Galvanic Skin Resistance
ECG	Electrocardiogram
EDA	Electrodermal Activity
SC	Skin Conductance
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
SCRAmp	Skin Conductance Response Amplitude
SCRPeaks	Skin Conductance Response Peaks Rate
SDNN	Standard Deviation of NN Intervals
HFN	Normalized High-Frequency Component
LF/HF ratio	(Low Frequency/High Frequency) Ratio
IS	Improvement Score
TCT-I	Improvement in Task Completion Time
CT-I	Improvement in Contact Time

Chapter 1 - Introduction

The rapid advancement and widespread adoption of virtual reality (VR) technology have opened new avenues for training and skill development across a broad range of industries, including healthcare, manufacturing, construction, and defence. As VR technology continues to evolve, it is imperative to understand its potential and limitations in enhancing the learning and performance of trainees in various skill-based tasks. This dissertation explored the effectiveness of immersive virtual reality (IVR) for industrial skills training by conducting a comprehensive review of the academic literature and industry practices as well as three controlled between-subjects experiments with 290 participants across them. The first chapter introduces the research, outlining the motivation, background, research questions, objectives, and scope of the study, setting the stage for a detailed investigation into the role of IVR in skills training and its potential implications for industry applications.

1.1.1. Virtual reality

Virtual Reality (VR) refers to computer-generated, immersive environments that simulate three-dimensional worlds, allowing users to interact with and explore virtual objects and surroundings in real-time (Jerald, 2016). It relies on a combination of hardware, for example. head-mounted displays (HMDs) as well as software to create a sense of presence and immersion, blurring the line between the physical and the virtual world (Cummings & Bailenson, 2016). VR has been applied in various fields, including education, entertainment, medicine, and training, transforming the way people learn, work, and interact with digital content. As VR technology continues to advance, it offers new opportunities for enhancing training effectiveness and efficiency, particularly in skill-based tasks that require realistic simulations and hands-on practice.

VR has a rich history, with its origins dating back to the 1960s when Morton Heilig invented the Sensorama (Heilig, 1962) as seen in Fig. 1 and Ivan Sutherland developed the first HMD system, known as the Sword of Damocles (Sutherland, 1968). Since then, VR has evolved significantly, with major advancements in both hardware and software technologies. In the 1980s and 1990s, VR captured the public's imagination with pioneering VR systems like NASA's Virtual Interface Environment Workstation (VIEW) (Fisher et al., 1988) and Virtuality arcade systems[5]. VR technologies can be placed within the context of Milgram's reality-

[5] https://web.archive.org/web/20200807083857/https://virtuality.com/

virtuality continuum, a spectrum ranging from authentic environments to fully immersive virtual environments (Milgram & Kishino, 1994). Along this continuum, different levels of immersion and interaction can be achieved, depending on the combination of real and virtual elements.

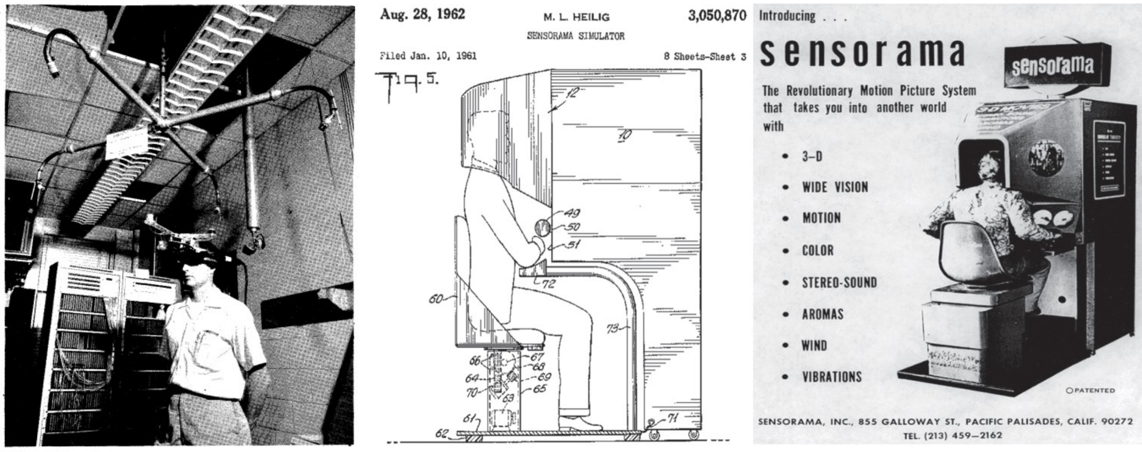


Fig. 1. The first VR system developed by Ivan Sutherland (Sutherland, 1968). (a) Illustration from the Sensorama patent (US 3050870) by Morton Heilig. (b) The product advertisement for the Sensorama from the 1960s (Cameron, 2017).

Augmented reality (AR) is another technology on Milgram's continuum (Milgram et al., 1995) (Fig. 2), situated between the fully real and fully virtual environments. AR integrates digital information and objects into the user's perception of the real world, offering a different set of applications and benefits compared to VR (Azuma, 1997).

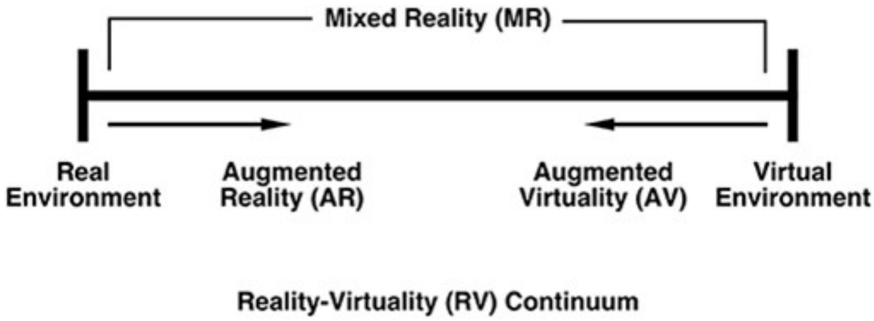


Fig. 2. Milgram's reality-virtuality continuum (Skarbez et al., 2021)

Today's VR systems can be broadly categorised into *non-immersive*, *semi-immersive* and *fully immersive* experiences. Non-immersive VR typically involves the use of desktop displays, while semi-immersive systems incorporate large projection screens and motion tracking devices. Fully immersive VR experiences are achieved through the use of HMDs, CAVE (CAVE Automatic Environments), haptic devices and advanced motion tracking systems (Sherman & Craig, 2003).

The current state of the art in VR technology includes high-resolution displays, low-latency motion tracking and sophisticated haptic feedback systems that provide users with an unprecedented level of immersion and interactivity. The applications of VR extend far beyond industrial scenarios, with a wide range of use cases found in rehabilitation, education, and therapy. For instance, VR has been used for physical rehabilitation, allowing patients to practice motor skills in a controlled and engaging environment (Holden, 2005). In the educational context, VR has been shown to improve students' engagement and learning outcomes in various subjects ranging from science and history to language and art (Merchant et al., 2014). Furthermore, VR has demonstrated therapeutic potential for the treatment of various psychological disorders such as phobias, post-traumatic stress disorder and anxiety (Riva et al., 2007).

1.1.2. The market for VR technologies

According to GlobalData (2022), the VR market is expected to grow from USD 6.9 billion in 2021 to USD 51.5 billion by 2030 at a CAGR of 25.1%. The increasing adoption of VR technology across verticals including gaming and entertainment, healthcare, automobile, architecture, and education is fuelling the growth of the market. Interestingly, this growth in VR adoption was not obvious just a few decades back, as the initial excitement around VR in the 80s wore off due to several factors, including technical limitations due to low-resolution displays, high latency, and cumbersome hardware, which led to less-than-optimal user experiences (Bailenson, 2018). The high cost of VR systems was another factor, which made them inaccessible to the general public and limited their practical applications (Burdea & Coiffet, 2003).

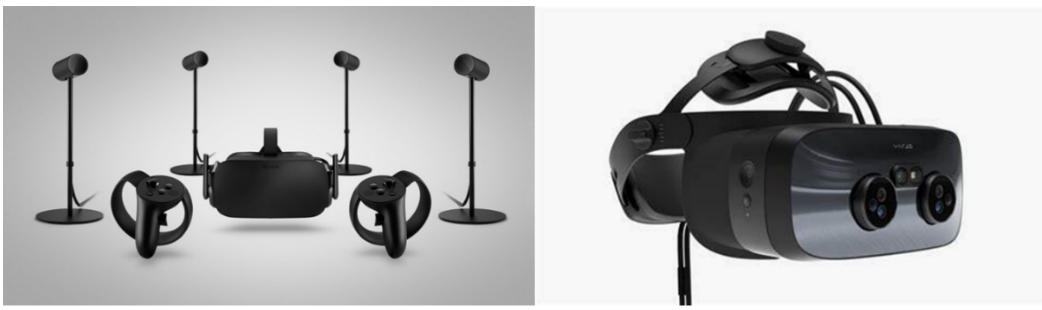


Fig. 3. (Left) Oculus Rift CV1, among the first VR headsets to undergo mass adoption. Note the base stations surrounding the headset used for tracking the position of the headset and the two controllers. (Right) The Varjo XR-3 headset with inside-out tracking and both VR/AR (passthrough) modes.

The 2010s marked a revival of interest in VR, catalysed by the invention of the Oculus Rift (Fig. 3). The successful Kickstarter campaign for Oculus Rift and its subsequent acquisition by Facebook in 2014 demonstrated the potential for VR to become a mainstream technology. Technological advancements such as higher-resolution displays, lower latency and more compact hardware enabled improved VR experiences (Cummings & Bailenson, 2016). The resurgence of VR sparked interest in various industries, including healthcare (Polce et al., 2020), education (Radianti et al., 2020) and industrial skills training (Abich et al., 2021). This renewed interest in VR has driven researchers and practitioners to explore its potential for revolutionising traditional training methods.

1.1.3 VR for industrial skills training

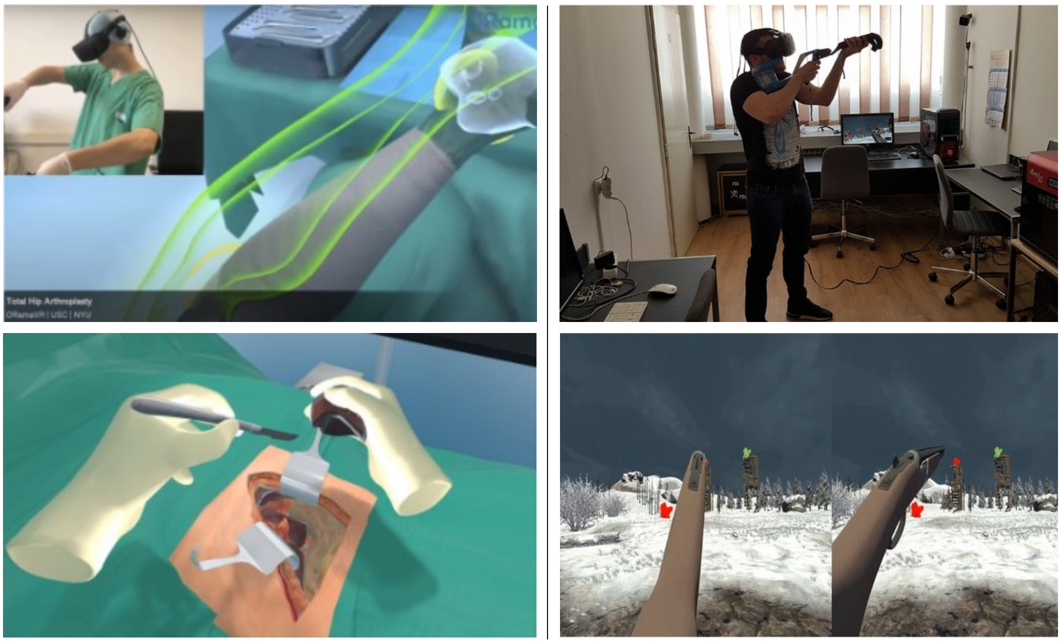


Fig. 4. Different configurations of VR use for training. (Top left) Total hip arthroplasty training using the OramaVR platform. (Bottom Left) Hemihepatectomy (Liver) surgical simulation using OramaVR (Zikas et al., 2023). (Right) VR shooting trainer with custom 3D-printed attachments for VR controllers (Harvey et al., 2019).

In the context of industrial skills training, VR offers numerous advantages over traditional methods, such as enhanced safety, reduced costs, and greater flexibility. For instance, VR-based training can provide a risk-free environment for trainees to practice complex tasks, thereby minimising the risk of accidents or damages (Chittaro & Buttussi, 2015). Additionally, VR training can be easily customised to suit the specific needs of different industries (see Fig.4), enabling the development of tailored training solutions that enhance skill acquisition and retention (Alaker et al., 2016). Furthermore, recent advancements in VR technology, such as immersive headsets, haptic feedback devices and sophisticated motion tracking systems, have significantly improved the realism and effectiveness of VR-based training experiences (Bailenson, 2018). VR has been successfully implemented in various industrial domains (see

Fig.5), including medical training (Jain et al., 2020; Mariani et al., 2021), maintenance and assembly procedures in manufacturing (Winther et al., 2020) and construction safety training (Kassem et al., 2017).

Fig. 5. Examples of VR training. (Top) VR simulation to train pump maintenance (Winther et al., 2020). (Bottom) VR assembly task simulation (Koumaditis et al., 2020b).

1.2 Theoretical background

1.2.1 Principles of motor skills learning

Magill and Anderson (2016) define motor skills as “activities or tasks that require voluntary control over movements of the joints and body segments to achieve a goal”. Motor skills can be broadly categorised into two types: *fine motor skills*, which involve precise, small muscle movements (e.g., writing or buttoning a shirt), and *gross motor skills*, which involve larger muscle movements and whole-body coordination (e.g., walking or throwing a ball). *Motor skills training* refers to the process of acquiring, refining, and maintaining motor skills through practice and experience. Motor skills training often involves a combination of instruction, demonstration, feedback and repetition, with the goal of improving performance, increasing automaticity and reducing errors (Wulf et al., 2010).

From a neuroscientific perspective, when a new motor skill is learned, several changes occur in the brain. These changes are primarily associated with neural plasticity, which is the brain’s ability to reorganise and adapt its structure and function in response to experience (Pascual-Leone et al., 2005). As individuals practice a new motor skill, synaptic connections between neurons are strengthened, and the efficiency of neural pathways involved in the skill increases (Kleim & Jones, 2008). Moreover, learning a new motor skill leads to the formation of motor representations in the primary motor cortex, which is responsible for encoding the spatial and temporal aspects of the skill (Dayan & Cohen, 2011).

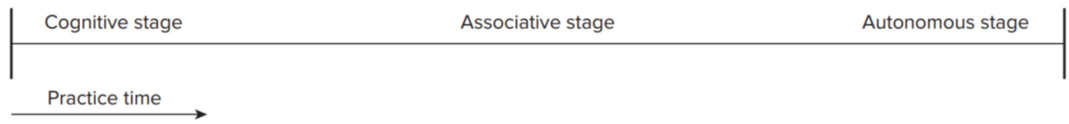


Fig. 6. The three stages of motor skills training over time according to Fitts and Posner's model. Source:(Magill & Anderson, 2016).

Fitts and Posner's stages of motor learning offer a framework for understanding motor skill acquisition, progressing from the *cognitive* stage to the *associative* stage and ultimately to the *autonomous* stage (see Fig. 6). This model has informed instructional methods and training protocols in various domains, such as sports, rehabilitation, and industrial skills training. The stages represent the progression from initial skill acquisition to automaticity, detailing the changes in cognitive processing and performance characteristics as a motor skill is learned. In the cognitive stage, learners develop a mental representation of the task, relying on verbal instructions, demonstrations, and feedback. The associative stage involves refining motor control and focusing on the finer details of the skill (Wulf et al., 2010). Finally, the autonomous stage is characterized by the ability to perform the skill with minimal conscious effort, allowing the learner to focus on higher-level cognitive processes or attend to other tasks simultaneously. It should be noted that the progression between these stages is gradual and not abrupt.

Fitts and Posner point out that reaching the autonomous stage is not guaranteed for every motor skill acquired. The amount/quality of practice and the accompanying instruction are important for learners to be able to reach this stage (Magill & Anderson, 2016). While Fitts and Posner provide a theoretical framework for motor skills learning, subsequent research in the movement sciences, medicine and rehabilitation has focused on the factors that enhance motor skills training. In their review of motor skills training literature from the fields of psychology and movement sciences with implications for medical training, Wulf et al. (2010) discuss factors that have been shown to enhance the learning of motor skills. The review highlights four factors: observational practice, the learner's focus of attention, feedback, and self-controlled practice. *Observational practice*, including dyad practice where two learners alternate between physical practice and observation of the other's practice, has been found to make important contributions to learning. Regarding the learner's focus of attention, instructions inducing an *external focus* (directed at the movement effect) are more effective than those promoting an *internal focus* (directed at the performer's body movements). This is because of the fact that directing attention towards the movement effect (external focus) promotes the use of automatic or unconscious processes, while an internal focus on one's own movements results in a more conscious form of control that limits the motor control system. Feedback was observed to not only have an informational function but also motivational properties, and self-controlled practice was found to be more effective than externally controlled practice conditions.

While Wulf et al. (2010) focused on factors that enhance motor skills learning in medical training, Sattelmayer et al. (2016) reviewed literature related to motor learning principles in general and identified four principles that impacted motor learning in different contexts. Specifically, the four principles identified by Sattelmayer et al. were: part practice or whole practice, random practice or blocked practice, mental practice, and augmented feedback. *Part practice* involves breaking a procedural skill into fundamental movement segments before combining them, while whole practice involves teaching the entire procedure as a whole entity. *Random practice* involves practicing multiple components of a procedural skill in a random order, while *blocked practice* requires skills to be practiced in closed blocks. *Mental practice* involves learning a procedure without physically performing it, using exercises such as thinking about the procedure and using imagery techniques. *Augmented feedback* provides information about performance that supplements sensory feedback from an external source, such as an educator or computer. In assessing learning, post-acquisition tests measure performance immediately after learning, retention tests measure performance after a rest period to eliminate temporary effects of the intervention, and transfer tests measure the ability to adapt a newly learned skill to a different situation, indicating the degree of learning.

1.2.2 Factors affecting learning in VR

The Cognitive Affective Model of Immersive Learning (CAMIL) (Makransky & Petersen, 2021) provides a research-based theoretical framework for understanding the learning process in IVR. The CAMIL proposes that the instructional methods used in IVR can be more effective when they facilitate the unique affordances of the medium, such as immersion, control factors and representational fidelity. Several studies have shown that some instructional methods, for example pre-training and generative learning strategies, can be more effective when presented in IVR compared to video or desktop VR (Meyer et al., 2019). The theory also suggests that the affordances of IVR, among these presence and agency, can render certain instructional methods more effective than when presented through non-IVR media, for instance the embodiment principle, which states that learners can benefit from human-like gestures and facial expressions in onscreen agents. The CAMIL predicts that there will be an interaction between media and methods, with learners in IVR-based lessons benefiting more from certain instructional methods than learners in video-based lessons. The model also identifies six affective and cognitive factors that lead to knowledge acquisition and transfer in IVR-based learning: interest, motivation, self-efficacy, embodiment, cognitive load and self-regulation. These theoretical underpinnings help explain how VR as a medium can impact learning and training by offering immersive, interactive and engaging experiences that support various cognitive and motivational processes. By leveraging these theories, researchers and practitioners can design more effective VR-based learning and training interventions.

1.2.3 The need for improving effectiveness of immersive VR training

The potential of IVR as a transformative tool for industrial skills training across various domains, including healthcare, manufacturing, and construction, has garnered increasing attention. While research has demonstrated the effectiveness of IVR for training procedural, decision-making, spatial, and fine/gross motor skills, a comprehensive understanding of the underlying factors that contribute to its success remains limited, even as the need for it in the industry is ever present. Moreover, the exploration of biosensors and haptic feedback technologies in IVR training, as well as the potential for remote training using current IVR technologies, has not been extensively investigated.

To address these gaps in the current body of knowledge, there is a pressing need to explore the effectiveness of IVR for industrial skills training and identify the critical parameters that influence its success. The research questions in the following section are formulated to guide the investigation and contribute to a more comprehensive understanding of IVR's role in skill development and its implications for industry applications. Addressing these research questions will not only help refine existing IVR training systems but also inform the design and implementation of novel, effective IVR training solutions across various industries.

1.3 Research questions

The increasing adoption of VR technology for various applications has sparked interest in exploring its potential for skills training across diverse domains. IVR provides an engaging and safe environment for learning and practicing complex tasks without the risks and costs associated with real-world training. Despite the promising outlook, several questions remain unanswered regarding the effectiveness of IVR for skills training and the factors influencing its success. This research aims to address these questions and enhance our understanding of IVR's role in training.

- *RQ1: What is the current state of the art in academic literature and industry practise regarding skills training using IVR?*

  A comprehensive understanding of the current state of the art in both academic literature and industry practice is essential to identify gaps and opportunities for further research in skills training using IVR. This research question seeks to establish a foundation for the study by reviewing existing knowledge and exploring real-world applications of IVR in skills training.

- *RQ2: Is IVR training effective compared to physical training?*

  To validate the efficacy of IVR training, it is crucial to compare its effectiveness with traditional physical training methods. This research question aims to investigate the relative merits of IVR training and determine if it offers the same advantages as physical training in terms of learning outcomes and skill acquisition.

- *RQ3: What is the link between the physiological arousal level of the trainees and the effectiveness of IVR training?*

  Physiological arousal, as a reflection of trainees' emotional and cognitive states, could influence the effectiveness of IVR training. Understanding the link between physiological arousal and the effectiveness of IVR training can help optimise the design of training scenarios and maximise learning outcomes.

- *RQ4: Can haptic feedback make IVR training more effective?*

  Haptic feedback plays a critical role in facilitating the learning of motor skills and enhancing the sense of presence in IVR environments. Investigating the link between different haptic feedback modalities and IVR training effectiveness can help identify the most suitable feedback mechanisms for specific training tasks and improve the overall quality of IVR-based training.

- *RQ5: What is the link between adaptive training and the effectiveness of IVR training?*

    Adaptive training, which tailors the learning experience to individual needs and performance, has the potential to enhance the effectiveness of IVR training. This research question seeks to explore the link between adaptive training and IVR effectiveness, providing insights into the benefits of personalised learning approaches in virtual environments.

By addressing these research questions, this study aims to contribute to the body of knowledge on IVR for skills training and provide valuable insights for the design and implementation of effective IVR training systems across various industries.

1.4 Methodology overview

1.4.1 Research methodology

This section will explore the importance of comprehending the current state of knowledge in both industry and academia through a systematic literature survey and industry case studies. By examining existing research and real-world applications, this dissertation primarily aims to identify the most effective and innovative practices in IVR training for industrial skills development. Furthermore, this comprehensive understanding will serve as a foundation for the design and execution of controlled experiments.

1.4.1.1 Systematic literature review

Systematic literature reviews are a rigorous and structured approach to reviewing existing literature on a specific research topic or question. This methodology involves the identification, selection and synthesis of high-quality research evidence to provide a thorough and unbiased overview of the current state of knowledge in a particular field (Petticrew & Roberts, 2008). Systematic reviews aim to minimise bias and ensure transparency and reproducibility by following predefined protocols, including explicit inclusion and exclusion criteria, search strategies and data extraction methods (Higgins et al., 2019). One of the main advantages of systematic literature reviews is their ability to identify gaps in the existing knowledge, helping researchers to formulate new research questions or hypotheses that can advance the field. Furthermore, by synthesising the findings from multiple studies, systematic reviews can provide more reliable and generalisable conclusions than individual studies, thereby supporting evidence-based decision-making in various domains, including industrial skills training.

In the context of this dissertation, a systematic literature review was conducted to explore the current state of the academic literature on industrial skills training across various industry domains. The intention is to establish a foundation for the experiments detailed in the subsequent chapters, which will focus on the use of VR for skills training and the factors that influence its effectiveness.

1.4.1.2 Case studies

Case studies are qualitative research methods that involve in-depth investigations of a particular phenomenon, issue, or context within its real-world setting. By focusing on a small number of cases, this methodology allows researchers to explore the complexities and nuances of the subject matter, providing rich and detailed insights that may not be achievable through other research methods (Stake, 1995). Case studies often employ multiple sources of data, such as interviews, observations, documents and artifacts, which can be triangulated to enhance the validity and credibility of the findings (Baxter & Jack, 2008). In the field of industrial skills training, case studies can provide valuable insights into the real-world applications of VR technologies, the challenges faced by organisations in implementing these technologies, and the factors that contribute to their success or failure.

In this dissertation, three case studies detailing the use of VR for skills training in Denmark will be presented, offering an opportunity to examine the practical applications of VR-based training in various industrial contexts. These case studies will complement the findings from the systematic literature review and thus provide an extensive understanding of the potential and challenges of using VR for industrial skills training as well as inspiration for empirical investigations for the dissertation.

1.4.1.3 Pilot studies

Pilot studies are small-scale, preliminary research studies conducted to evaluate the feasibility, time, cost, risk and potential effectiveness of a research design or methodology before implementing it on a larger scale (Leon et al., 2011). They are often used to refine and optimise various aspects of a controlled experiment, for instance the research protocol, data collection methods, recruitment procedures and intervention components (Thabane et al., 2010). Pilot studies help in designing controlled experiments by:

- Assessing the feasibility of the research design: By testing the research design on a smaller scale, pilot studies can provide valuable insights into its feasibility, allowing researchers to make any necessary modifications to improve its viability and effectiveness.
- Refining the experimental protocol: Pilot studies can help researchers fine-tune their experimental procedures, such as the timing of interventions, the wording of instructions or the administration of questionnaires, ensuring that the main experiment runs smoothly and efficiently.
- Evaluating data collection methods and instruments: Pilot studies allow researchers to test the reliability and validity of their data collection tools, including questionnaires, interviews or physiological measurements, and make any necessary adjustments to ensure that they accurately capture the variables of interest.

1.4.1.4 Controlled experiments

Controlled experiments help investigate questions related to effectiveness of VR in skills training. They can also contribute to establishing internal and external validity. *Internal validity* refers to the extent to which a study establishes a trustworthy cause-and-effect relationship between the manipulated independent variable (e.g., VR training) and the measured dependent variable (e.g., improvement in performance). A study with high internal validity allows researchers to confidently conclude that changes in the dependent variable are caused by the independent variable, rather than by confounding factors or biases. Controlled experiments can enhance internal validity by:

- Controlling for confounding factors: By holding constant or controlling for factors that might influence the outcome, researchers can minimise the risk of confounding variables affecting the results.
- Ensuring random assignment: By randomly assigning participants to different experimental conditions (also known as a between-subjects method), researchers can reduce the impact of individual differences on the results and ensure that the groups are comparable.
- Conducting pre- and post-test measurements: By measuring the dependent variable both before and after the intervention, researchers can assess the changes that occur

because of the VR training and control for pre-existing differences in skills or knowledge.

External validity, on the other hand, refers to the extent to which the results of a study can be generalised to other contexts, populations or settings. A study with high external validity allows researchers to apply their findings beyond the specific conditions of the experiment.

1.4.2 Techniques and tools

1.4.2.1 Biosensors for detecting physiological arousal

Physiological arousal refers to the activation of the autonomic nervous system (ANS) in response to stimuli such as stress, excitement or fear (Cacioppo et al., 2007). The ANS is a part of the nervous system responsible for controlling involuntary bodily functions such as heart rate, respiration, and digestion. It has two primary subsystems: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The SNS is responsible for the fight-or-flight response, which prepares the body for action in response to perceived threats or stressors. When the SNS is activated, it increases the heart rate, the respiration, and the blood flow to the muscles, while inhibiting non-essential functions like digestion. The PNS is responsible for the rest-and-digest response, which conserves energy and helps the body recover from stress or activity. When the PNS is activated, it slows down the heart rate and respiration and promotes digestion, returning the body to a state of calm and relaxation. Electrodermal activity (EDA) and heart rate variability (HRV) are two commonly used measures of physiological arousal that reflect the activity of the ANS:

- *Electrodermal activity (EDA)*, also known as skin conductance, is a measure of the electrical conductance of the skin, which varies with its moisture level. It serves as an activity indicator of the sympathetic nervous system, reflecting emotional and cognitive states such as stress, arousal and attention (Boucsein, 2012). Research has shown that monitoring skin conductance can provide insights into the effectiveness of VR-based training interventions and participants' engagement levels (Katsis et al., 2008). The Shimmer GSR+[6] (see Fig. 7) was used to measure skin conductance in experiments conducted as part of this dissertation.

[6] https://shimmersensing.com/product/shimmer3-gsr-unit/

- *Heart rate variability (HRV)* refers to the variation in time between successive heartbeats. It is a widely used non-invasive measure of autonomic nervous system activity, with lower HRV values indicating higher stress levels and reduced adaptability (Shaffer & Ginsberg, 2017). By monitoring HRV, researchers might assess the impact of VR training on participants' stress and cognitive load, which can have implications for training effectiveness and retention. The Polar H10[7] electrocardiogram (see Fig.7) sensor was used to measure HRV in the experiments of this dissertation.

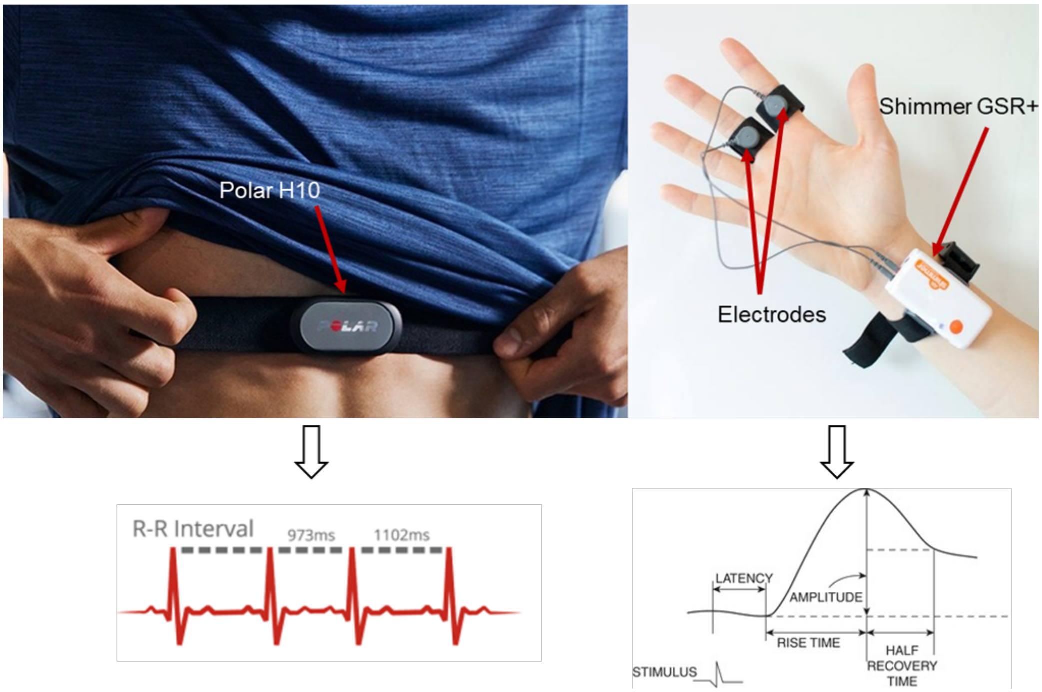


Fig. 7. (Left) The Polar H10 ECG sensor for measuring heart rate variability metrics. (Right) The Shimmer GSR sensor for measuring electrodermal activity.

1.4.2.3 Software tools

Unity – the Unity game engine[8] is a powerful and versatile platform for creating interactive 3D and 2D content, including VR applications (Goldstone, 2017). Its flexibility and ease of use has made it a popular choice for developing VR-based training scenarios, allowing researchers to create realistic and immersive environments tailored to the specific requirements of their studies (Erickson et al., 2019).

[7] https://www.polar.com/us-en/sensors/h10-heart-rate-sensor/
[8] https://unity.com/

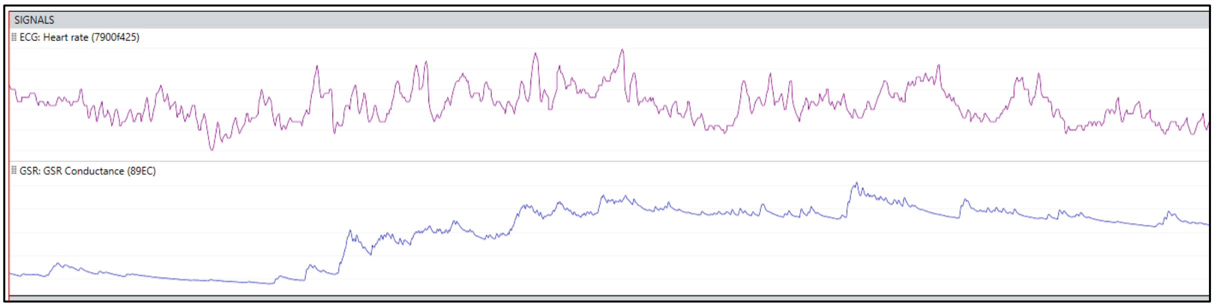


Fig. 8. Sample heart rate and skin conductance signals of a study participant.

iMotions – the iMotions[9] biometrics research platform is a comprehensive solution for collecting, synchronising, and analysing multiple biometric data streams, including skin conductance and HRV. See Fig. 8 for a sample signal recording of a study participant's heart rate and skin conductance values. This platform was used in this dissertation as a data store for performance and physiological arousal data.

1.4.2.4 VR Head Mounted Displays (HMDs)

In this research, three different VR HMDs were used to conduct the experiments: the Oculus Rift, Oculus Quest 1, and Oculus Quest 2. These devices have been widely adopted in various VR applications due to their high-quality performance, user-friendly features, and relative cost-effectiveness.

- Oculus Rift: Launched in 2016, the Oculus Rift is a tethered VR headset that requires connection to a PC for operation. It features a high-resolution display, a refresh rate of 90 Hz and a wide field of view (FOV) of around 100 degrees, providing an immersive VR experience. The Rift uses an external tracking system, known as the Oculus Constellation, for accurate positional tracking of the headset and the Oculus Touch controllers, which enable precise hand and finger movements within the virtual environment (Oculus VR, 2016).
- Oculus Quest 1: Released in 2019, the Oculus Quest 1 is a standalone VR headset that does not require a connection to a PC. It features a high-resolution display, a refresh rate of 72 Hz and a slightly narrower FOV compared to the Rift. The Quest 1 incorporates a built-in inside-out tracking system, which eliminates the need for

[9] https://imotions.com/

external sensors. The device uses the same Oculus Touch controllers as the Rift, ensuring intuitive and accurate hand interactions in VR (Oculus VR, 2019).

- Oculus Quest 2: Launched in 2020, the Oculus Quest 2 is an upgraded version of the Quest 1, offering improved performance and features. It boasts a higher-resolution display and an adjustable refresh rate of up to 120 Hz (with a recent software update), enhancing the visual fidelity and smoothness of the VR experience. Like the Quest 1, the Quest 2 uses an inside-out tracking system and is compatible with the Oculus Touch controllers. Additionally, the device can be connected to a PC via the Oculus Link cable, providing access to PC VR content, and increased graphical capabilities (Oculus VR, 2020).

1.4.2.5 Haptic feedback

Haptic feedback refers to the technology that provides tactile sensations to users through vibrations or forces, simulating the sense of touch. The different types of haptic feedback (Culbertson et al., 2018) commonly available in commercial hardware are:

- *Vibration feedback*: This type of haptic feedback uses motors or actuators to generate vibrations, stimulating the sense of touch through oscillatory motion. It is commonly used in consumer electronics, such as smartphones, smartwatches, and gaming controllers. In this dissertation, for some of the experiments, the vibration feedback available in the Oculus Quest controllers (Fig. 9) are used for giving feedback to participants. We also developed custom devices to provide vibrotactile feedback for the experiment detailed in chapter 4.
- *Force feedback*: Force feedback systems exert forces on users, simulating the feeling of touch or resistance when interacting with virtual objects. This type of haptic feedback is often used in more advanced applications, such as surgical simulations, virtual sculpting, or training scenarios. The Geomagic Touch[10] (Fig. 9) was used in experiment detailed in chapter 4. It is commonly interfaced to a virtual scene, where a user holding the device handle can move and orient a digital probe in 3D space. The device provides coherent force feedback along the three orthogonal directions when the probe interacts with virtual objects.

[10] https://www.3dsystems.com/haptics-devices/touch

- *Tactile feedback*: Tactile feedback focuses on the cutaneous area using various technologies, such as (but not only) electrostatic or pneumatic actuators, to create surface texture or shape changes that mimic the feeling of touch. This type of haptic feedback is used in applications like wearable devices or touchscreens.

Figure 9 Two common types of haptic feedback devices. On the left is a Geomagic Touch which recreates directional force feedback the user's hands when interacting with virtual objects. On the right is an Oculus Quest 2 controller which provides vibrations to the user's fingers and palm depending on their actions inside the virtual environment.

1.5 Dissertation structure

The following outlines the structure and content of the subsequent chapters and provides a brief overview of each chapter and its contribution to the overall dissertation.

1.5.1 Chapter 2: Systematic literature review and industry case studies on IVR based industrial skills training

1.5.1.1 Systematic literature review

The systematic literature review analysed 78 representative studies (see Fig.10 for the process) to answer three key questions: Is IVR an effective training method for industrial skills training? How is research in this field applied? How can we make IVR training more effective and applicable for remote training? The results showed that IVR is a promising training method with high effectiveness scores. The analysis also revealed several gaps in the application of IVR training, for instance a lack of learning theories in the design process and limited metrics beyond time and scores. The study exposed unexplored avenues of research, including the utilisation of biosensors for data collection, haptics that increases realism and applications with remote training potential. The research was published in the journal Behaviour and Information Technology under the title "A Systematic Review of Immersive Virtual Reality for Industrial Skills Training" (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021).

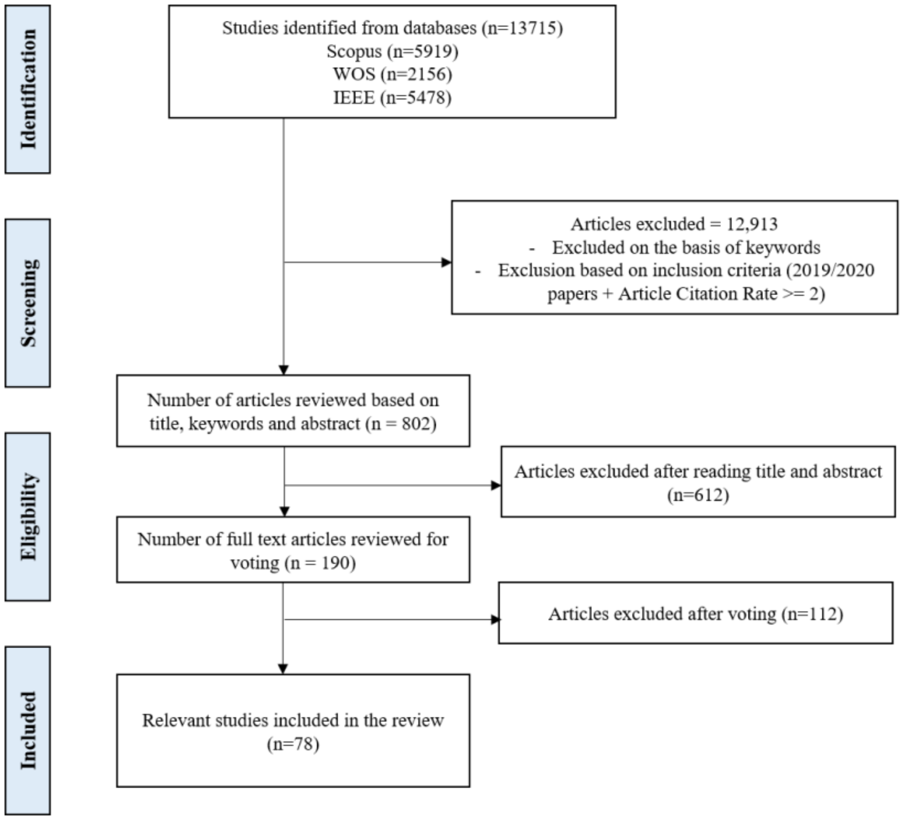


Fig. 10. The systematic literature review process.

1.5.1.2 Industry case study

The case study depicts three cases from the Danish industry (Fig. 11) to showcase the motivation, technology, design, and perception of IVR adoption. The research questions addressed here include the motivation for IVR adoption, the technological and design elements incorporated in IVR training and the assessment and perception of IVR training by stakeholders. The methodology of the research involves a thorough inspection of each case, including semi-structured interviews, inspection of the IVR application and field studies, and analysis of open published data and available reports. The three cases described are Siemens Gamesa, a Spanish-German wind engineering company, DSB, the largest Danish train operating company, and Grundfos, the largest pump manufacturer in the world. This research was published as a paper proceeding in IEEE VR 2021 and was titled "Immersive Virtual Reality Training: Three Cases from the Danish Industry" (U. Radhakrishnan et al., 2021).

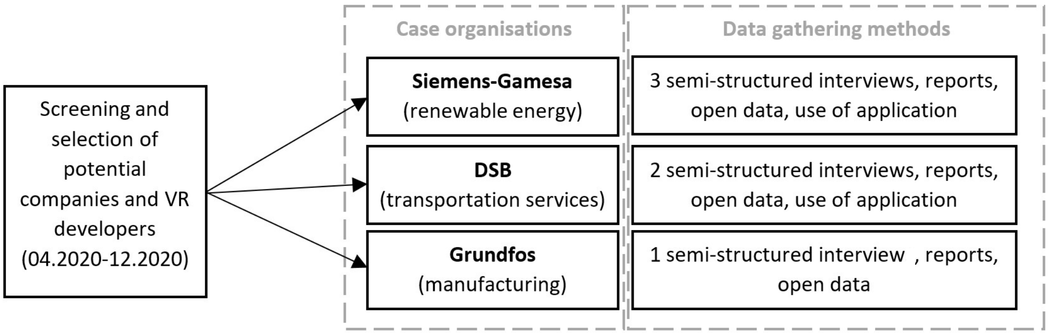


Fig. 11. Case studies on the use of IVR in the Danish industry.

1.5.2 Chapter 3: Investigation of Effectiveness of VR for fine motor skills training and the link to physiological arousal

This chapter details an experiment which investigates the effectiveness of IVR for training participants in a fine motor skill task (buzz-wire) and its link to physiological arousal. The results showed that VR training (Fig. 12) is as good as or even slightly better than physical training in improving task performance, and participants trained using IVR reported an increase in self-efficacy and immersion. The study highlights the potential of using arousal and training performance data for designing adaptive VR training systems. This chapter has been published in the Virtual Reality journal under the title: *"Investigating the Effectiveness of Immersive VR Skill Training and its Link to Physiological Arousal"* (Radhakrishnan, Chinello, et al., 2022).

Fig. 12. (Left) VR setup for the motor skills training experiment. (Right) Physical setup.

1.5.3 Chapter 4: Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task

This chapter explores the connection between fine motor skill training in VR (see Fig. 13), haptic feedback, and physiological arousal. The experiment, involving 73 participants, compared the effects of three feedback conditions: visual/kinesthetic, visual/vibrotactile, and visual only (i.e., no haptic feedback). Results showed performance improvements across all conditions but no change in self-efficacy, presence, or task load. The visual/kinesthetic feedback condition led to higher physiological arousal compared to the visual/vibrotactile condition, and higher arousal levels correlated with better performance. The findings suggest that haptic feedback can influence arousal levels, encouraging further research on its potential for enhancing motor skill training in VR. This chapter is under journal review and is titled *"Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task"*.

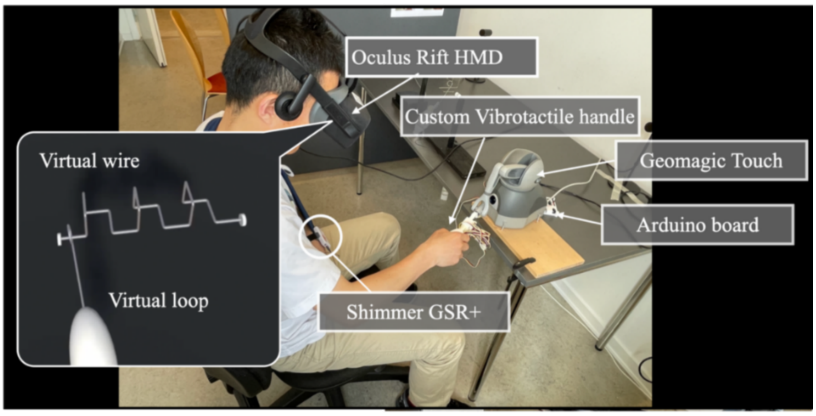


Figure 13. Setup for the haptics experiment with a Geomagic Touch (for kinesthetic feedback) and a custom handle for vibrotactile feedback during training.

1.5.4 Chapter 5: A Controlled, Preregistered Experiment on Self-Efficacy and Performance in Adaptive Virtual Training

This chapter examines the effectiveness of adaptive versus fixed training in an immersive virtual reality (IVR) setting (see Fig. 14) for fine motor skill development. Participants (N = 130) were randomly assigned to adaptive or fixed training groups. The results showed no significant differences between the groups in terms of performance or self-efficacy, indicating that more research is needed to determine when adaptive training is beneficial. Overall, training improved participants' accuracy and speed in a virtual test, but transfer of skill to real-world tasks showed mixed results, with increased accuracy but reduced speed. This chapter is an unpublished draft titled *"Training, Quickly and Accurately: A Controlled, Preregistered Experiment on Self-Efficacy and Performance in Adaptive Virtual Training"*.

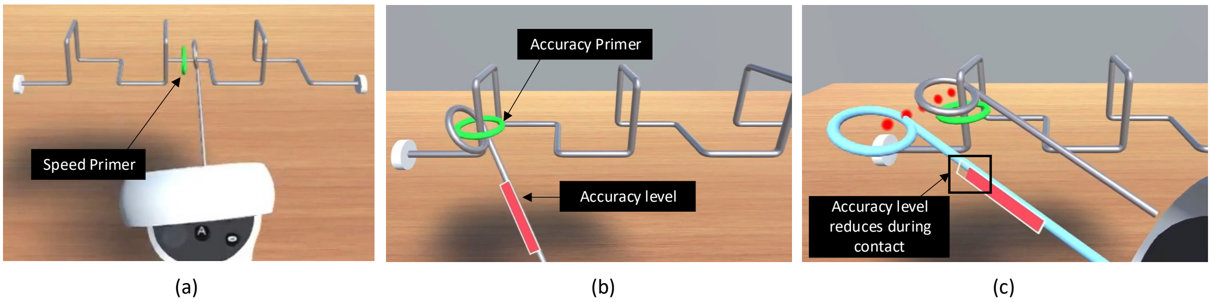


Figure 14 Screenshots from the adaptive IVR training experiment: (a) speed focused training, (b) & (c) accuracy focused training.

1.5.5 Chapter 6: Discussion and conclusion

This concluding chapter offers a comprehensive discussion of the research conducted throughout this dissertation, emphasizing its key contributions to the field. Additionally, it presents a critical examination of the study's limitations and identifies potential areas for future research and exploration.

Chapter 2 - Systematic literature review and industry case studies on VR-based industrial skills training

This chapter depicts the current state of Immersive Virtual Reality (IVR) in the context of industrial skills training by presenting a two-fold analysis. The first subsection provides a comprehensive systematic review of the academic literature, highlighting the use of IVR in industrial skills training, as well as the current understanding of its effectiveness and the research gaps that exist. The second subsection offers a closer look at real-world applications by discussing three industry case studies from the Danish market. These case studies showcase the motivation behind adopting IVR for industrial training, the technological and design characteristics implemented, and the stakeholders' perceptions of its applicability. By integrating both academic and industry perspectives, this chapter aims to present a well-rounded understanding of the current landscape and potential future developments in IVR for industrial skills training.

2.1 A Systematic Review of Immersive Virtual Reality for Industrial Skills Training

Virtual reality (VR) training offers the capability to industrial workers to acquire skills and address complex tasks by immersing them in a safe and controlled virtual environment. Immersive VR (IVR) training is adopted in many diverse settings, yet little systematic work currently exists on how researchers have applied it for industrial skills training and if it holds the potential to be applied remotely. In this review, 78 representative studies were analysed to answer three key questions: Is IVR an effective training method for industrial skills training? How is research in this field applied? And how can we make IVR training more effective and applicable for remote training? We can testify that IVR is a promising training method with high effectiveness scores. However, our analysis has uncovered several gaps in the application of IVR training, like the lack of learning theories in the design process and limited metrics beyond time and scores. Additionally, our review also exposed unexplored but intriguing avenues of research, like the utilisation of biosensors for users' data collection, haptics that increases realism and applications with remote training potential.

2.1.1 Introduction

Effective industrial training has always been paramount. Beyond the undeniable value of health and safety training, literature reports that the costs of fail-to-recall procedural tasks, for example in a production environment, are high and that errors can be catastrophic for the product and the overall production costs (A.-C. Falck et al., 2010). Similar shortcomings apply when errors occur beyond the production line, for example in the assembly of machinery or instalment of electronic components in the field of business or hard-to-access locations, such as offshore wind turbines or oil drilling rigs (U. Radhakrishnan et al., 2021; Reason & Hobbs, 2017). Still, providing effective training is not an easy task. The literature identifies variations in the way one can design, deliver and implement training programmes (Salas et al., 2012). On average, organisations spend 10% of their budget on learning tools and technologies, with the most popular ones being e-learning platforms, learning management systems and simulations (Freifeld, 2018). These investments in training activities allow organisations to adapt, compete, excel, innovate, produce, follow safety precautions, improve services and achieve business goals (Grossman & Salas, 2011).

As of late, a new innovative wave of interactive immersive virtual reality (IVR) systems is being utilised for industrial training (Radianti et al., 2020). The need for such implementations echoes the industrial requirements for cost-effective, safe, scalable, modular, and mobile systems and their potential to increase training effectiveness. From a virtual reality hardware perspective, this demand is covered by the availability of various head-mounted displays (HMDs), either low-budget for mobile devices, such as Google Cardboard and Samsung Gear VR, or high-end VR equipment like HTC VIVE and Oculus Rift. The availability of such devices and newcomers like the untethered Oculus Quest create an HMD market valued at USD 44.7 billion by 2024 with a compound annual growth rate (CAGR) of 33.5% during the forecast period (MarketsAndMarkets, 2019). A market that due to the COVID-19 restriction for remote work has the potential to mature at an even faster rate.

Research in IVR training systems depicts numerous benefits, including soft-skills acquisition (Daniel Eckert & Andrea Mower, 2020), increased engagement, presence and immersion (Buttussi & Chittaro, 2018; Jensen & Konradsen, 2018) and reduced cognitive load (Sun et al., 2019), to name a few. Therefore, it is not surprising that there is a growing interest from industry to invest in and academia to research this phenomenon. In terms of the latter, one may find several literature reviews spanning education (Jensen & Konradsen, 2018; Pellas et al., 2019; Radianti et al., 2020), serious games (Checa & Bustillo, 2019; Feng et al., 2018), adaptive systems (Zahabi & Abdul Razak, 2020), rehabilitation (Rose et al., 2018) and operator training simulators (Patle et al., 2019). While these reviews establish a solid ground for VR-based training, they do not provide a focus on industrial training. Thus, in this paper, we investigate the use of IVR in industrial settings and try to answer the following research questions (RQ):

RQ1 What types of skills does immersive VR training provide in the industry?

RQ2 Which learning theories are utilised in immersive VR industrial skills training?

RQ3 Which research designs, data collection methods and data analysis methods are utilised in immersive VR industrial skills training?

RQ4 What HMD technologies, biosensors and haptics are utilised in immersive VR industrial skills training?

RQ5 What levels of effectiveness of immersive VR are reported in industrial skills training?

RQ6 How applicable are current immersive VR applications to be provided as remote training solutions?

Herein, a systematic review of the existing VR research is presented with this section being the introduction, followed by section 2.1.2 that depicts the theoretical background, section 2.1.3 discusses previous systematic reviews, section 4 details the review process, section 2.1.5 explains the results, section 2.1.6 discusses the findings, section 2.1.7 lists future research directions and finally, section 2.1.8 summarizes the conclusions.

2.1.2 Theoretical background

In the following, we synthesise the three distinctive topics addressed in this review: immersive virtual reality, remote virtual training, and industrial skills training.

2.1.2.1 Immersive VR

Virtual reality, the artificially generated interactive digital environment designed to simulate real life, is mostly characterised by two attributes, namely 'immersion' and 'presence'. Herein with immersion, we refer to the "objective level of sensory fidelity a VR system provides" (Bowman & McMahan, 2007), whereas presence refers to the subjective experience of the user resulting from being in the immersive environment (Jensen & Konradsen, 2018). In most cases, immersive VR systems comprise robust tracking systems, head-mounted displays with in-built or external motion tracking sensors that grasp minute movements which refresh the visual stimuli close to real-time and thus provide deeper immersion. In an IVR experience, the user is frequently isolated from external visual cues and uncontrolled stimuli from his/her own physical world, allowing him/her to experience a highly engaging, interactive setting. Interactivity in such a setting allows the user to act, transform the IVR experience and interact with objects and tools in a desirable manner (Radianti et al., 2020).

2.1.2.2 Remote virtual training

Remote training refers to a training activity that is provided at a relative distance from a physical instructor and/or the main place of work (i.e., place and/or physical setting in which the trainee worker will apply his/her acquired skills). E-learning, online courses, webinars, and virtual training are common components of a remote training endeavour. In the case of virtual training, one may add digitisation and gamification of real-life scenarios, engaging and immersive digital content, human-computer interaction and, of course, immersive virtual reality. The latter is addressed in this paper.

VR and remote training are quite interweaved terms with overlapping characteristics (i.e., digital content, gamification, use of technology, etc.). However, it is our understanding that VR training is not always remote or at least holds the same level of mobility, either, due to the physical presence of a trainer, or grounded technological apparatus (i.e., haptic and positioning sensors) or both. Thus, as part of this research, we aim to expose the degree to which current immersive VR applications can be provided as remote training solutions and discuss the potential of IVR systems to become remote.

2.1.2.3 Industrial skills training

The term ‘skill’ refers to the ability to perform an action/task/job with determined results often within a given array of performance criteria like time, effort, etc. In many cases, skills are divided into domain-general and domain-specific skills. In this review, we adopt the term “industrial skills” to refer to a wide spectrum of skills required to perform one’s job in an industrial setting. However, our scope is to focus on a blend of the following four categories:

- *Perceptual motor or psychomotor skills* involve skills that require hand-eye coordination to solve the problem; wood carving and surgical skills are illustrative examples.

- *Procedural skills* refer to the learning of processes and sequences important in scenarios like learning safety and evacuation procedures or in the operation and assembly of machinery.
- *Decision-making or problem-solving skills* involve the skill of selecting one or a few alternatives out of many possible choices.
- *Spatial skill* is the capacity to understand, reason and remember the spatial relations among objects or space. Visual-spatial abilities in VR are required for navigating in the virtual environment, understanding or estimating distance and measurement, or understanding and placing 3D objects.

Industrial skills training is crucial to the economic development and competitiveness of nations (Tabbron & Yang, 1997). In a 2011 report by Deloitte and the Manufacturing Institute, it was found that among surveyed US manufacturers, a shortage of skilled manpower in machining, machine operation and crafting was negatively impacting the ability of manufacturers to expand operations and innovate (Morrison et al., 2011). They also observed that – as the nature of manufacturing jobs is rapidly changing, fuelled in part by automation – the skill levels of their current employees failed to catch up. In fact, in the same report, 74% of the surveyed companies felt that the shortage of skilled production workers had a negative impact, as compared to 19% for scientists and engineers. They further observed that a lack of problem-solving skills was the most serious skill deficiency among the respondent companies' workforces. To alleviate this, a new innovative wave of e-learning systems, among them interactive immersive VR solutions, is being used as training tools (Radianti et al., 2020). The need for such implementations echoes the industrial requirements for cost-effective, safe, scalable, modular and mobile systems as well as their current needs to increase training effectiveness.

2.1.3 Previous systematic literature reviews

Several, previous literature reviews hold merit and truly constitute the point of departure for our literature review. Examples of previous reviews include Radianti et al. (2020), Checa and Bustillo (2019), Feng et al. (2018), Jensen and Konradsen (2018), Suh and Prophet (2018) and Zahabi and Abdul Razak (2020). However, to contribute to theory, our study aims to fill the gaps in the existing literature. We intend to do so by placing the focus on the following two areas:

Immersive VR and industrial skills training – None of the above-mentioned reviews exclusively centred on IVR in an industrial setting. For example, Checa and Bustillo (2019) examine serious games and apply bibliographical, technical (typology) and evaluation (typology and methods) analysis, and only partially investigate training. Suh and Prophet (2018) identify several distinctive domains that use immersive technologies, including education, entertainment, healthcare and marketing; however, they do not map these in their analysis, neither do they focus exclusively on IVR. Other reviews deal with education and learning with an interesting but partial focus on industrial applications (Jensen & Konradsen, 2018; Radianti et al., 2020), and others investigate fascinating but narrow areas of interest like evacuation (Feng et al., 2018), operator training (Patle et al., 2019), or adaptive systems (Zahabi & Abdul Razak, 2020).

Use of haptics, biosensors and potential for remote training – In addition to the gaps described earlier, five reviews do not identify the typology and application of haptic systems and sensors in connection with VR training in industrial settings. One of the five, Suh and Prophet (2018), mentions the utilisation of haptic systems and sensors as technological stimuli, for sensory modality, perceptual stimuli and affective reaction, respectively; yet they do not proceed to a mapping or thorough analysis. On the other hand, Zahabi and Abdul Razak (2020) indeed describe the applicability of sensors in IVR, but mainly for physiological data retrieval and with no in-depth typology characterisation. Finally, none of the reviews focuses on the potential of the IVR applications to be applied as remote training solutions.

2.1.4 Review process

The filtering process was carried out following the Preferred Reporting Items for Systematic Literature Reviews and Meta-Analyses (PRISMA) framework (Moher et al., 2009). The papers were filtered following identified inclusion and exclusion criteria, which are listed in detail in the subsections below and depicted in Figure 15.

2.1.4.1 Search databases

We chose three prominent search databases with a strong focus on technology (Scopus, Web-of-Science and IEEE-Xplore) to build our literature review using the identified search terms. The decision was made after a thorough inspection and in recognition of a) the complexity (in format styles) that adding all the databases would cause and b) the fact that, in most cases, the data (publications) were repeated across the databases. Grey literature, like white papers from the industry, is absent from the review as only peer-reviewed academic publications were included. The search term in all three databases contained papers available online ranging from January 2010 to December 2020. We focused on papers after 2010 because immersive virtual reality research expanded greatly in the last decade, especially ever since the Oculus Rift and other commercial HMDs were released into the market. Similar literature reviews have timespans of the last decade, e.g. Suh and Prophet (2018) with 2010-17 and Jensen and Konradsen (2018) with a range of 2013-2017. And those that span beyond the last decade end up with representative papers mostly from the last ten years (Radianti et al., 2020; Zahabi & Abdul Razak, 2020).

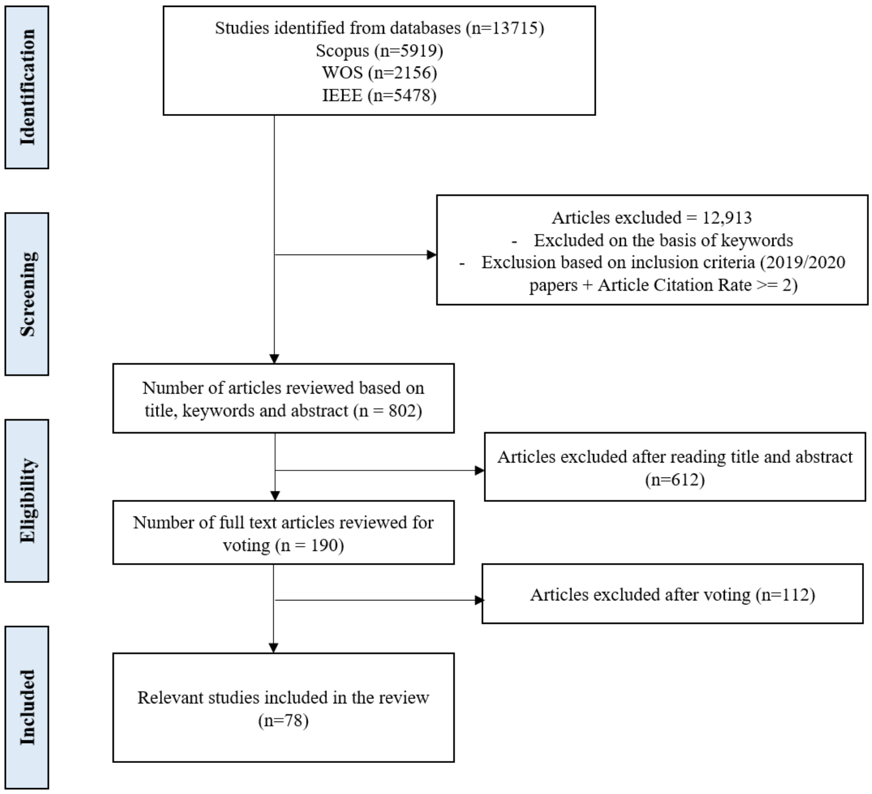


Figure 15 Literature review process

2.1.4.2 Search terms - Keywords

We formulated the following generic logical expression as the basis for the search strings across the three databases:

"VIRTUAL REALITY" OR "VR" AND "EDUCAT" OR "SKILL*" OR "LEARN*" OR "TEACH*" OR "TRAIN*" IN TITLES, ABSTRACTS AND PUBLICATION YEAR >=2010 AND PUBLICATION YEAR <=2020*

The star (*) symbol, when applied in the search terms, is the generic method to represent variations of the same term, for example, EDUCAT* represents the terms "educate", "education", "educator", "educating", etc. Furthermore, only the title and abstracts were included in the search parameters, as the keyword listings and abstracts can be quite generic and yield papers that have little relevance for the research questions being asked. The specific search expressions used in the three databases are listed in the Appendix. The three databases produced 13715 publications in total.

2.1.4.3 Automatic and semi-automatic filtering

Since the size of the combined database was substantial at 13715 publications, we employed automatic and semiautomatic filtering methods mentioned in the following sub sections to reduce the size of the resulting database.

2.1.4.3.1 Removing duplicates and incomplete entries

Duplicates were removed based on the publication title by using inbuilt functions in MS Excel. We also found several incomplete entries with most of the fields missing. These publications were checked in the original databases to ensure that legitimate entries were not discounted, and, after this, the remaining incomplete entries were removed.

2.1.4.3.2 Citation ranking

After duplicates that could be identified by an Excel search were removed, the database had 6420 unique papers. An automated filtering method was developed to identify high-quality papers in the database between 2010 and 2018, which filtered out those which did not have substantial citations in relation to their age (as they are not likely to be representative of the field). The Article Citation Rate (ACR) as suggested by Hutchins et al. (Hutchins et al., 2016) was calculated for each paper using the formula :

$$\mathrm{ACR} = \frac{\text{Total citations to article}}{\text{Last year in citation database} - \text{Year of article publication}}$$

The ACR gave a balanced weight to recent papers with fewer citations compared to papers with higher citations but old publication dates to help identify higher quality papers. We experimented with different ACR cut-off values and adopted a value of 2. For example, we found 673 papers with an ACR greater than or equal to 3, 1100 papers with an ACR greater than or equal to 2, and 1970 papers with an RI greater than or equal to 1. As an illustration, an older paper from 2011 with 5 citations would get an ACR score of 0.6 (5/9) which is lesser than

2 and thus rejected, while a paper with 25 citations from the same year would get a score of 2.8 (25/9) which would then be included. The possibility of papers missing the threshold by a few points does exist, but as these papers are likely to be marginal in their impact (in terms of citations as a proportion to their age), they are not likely to impact the purpose of the literature search. The 1100 papers with an ACR greater or equal to 2 between 2010 and 2018 was combined with all papers (regardless of their ACR score) from 2019 and 2020 resulting in 2439 papers that were fed into the semi-automatic filtering stage detailed in the next section.

2.1.4.3.3 Semi-automatic filtering based on exclusion criteria

The exclusion criteria were defined based on user type, application domain and technology, removing studies involving applications for users with mental disabilities (due to the narrow, specialised focus of these studies), non-industrial use cases and non-immersive VR. In more detail:

- **Exclusion criteria based on ability:** Parkinson’s, rehabilitation, autism, stroke, ADHD, multiple sclerosis, elderly, schizophrenia, Alzheimer’s, depression, Down’s syndrome, dyslexia, injury, PTSD, dementia, patient, deaf, stress, cybersickness, fatigue, motion sickness.
- **Exclusion criteria based on technology:** Papers which does not use any IVR at all but uses desktop VR, AR, CAVE, spherical/360 videos, machine learning not related to training, web-based VR. Papers comparing IVR to other VR technologies are not excluded.
- **Exclusion criteria based on domain/setting:** school, higher education, sports, soft skills, cybersecurity.
- **Exclusion criteria based on quality:** non-peer-reviewed articles.
- **Other exclusion terms:** child, science, elementary, museum, library, cultural heritage, art, architecture, tourism, language learning, foreign language learning, wireless LANs, soft skills, communication skills, public speaking, STEM, STEAM, special education, dance, remote/virtual labs, animal cognition, distance teaching.

2.1.4.3.4 Manual filtering

After the semi-automatic filtering stage was completed, 802 papers remained in the database for our consideration, providing a comprehensive body of literature, which then further filtered through manual processes (see Figure 15). First, a voting protocol was devised among the three authors, with a mark of '1' for acceptance and '0' for rejection based on the title, abstract, other bibliographic details and the main parts of the paper (if they chose to). This process was first tested and discussed on a sample of 100 papers. A Fleiss Kappa (Fleiss, 1971) inter-rater reliability score of 0.578 was calculated on the 802 ratings by three judges, showing moderate agreement (0.41-0.60). Papers with a score of three (i.e., all three authors agreed on their inclusion) were directly accepted for final review, while those with a score of two (i.e., where one of the authors disagreed) were marked for further detailed discussion, and only those with a score of one or zero were rejected. Following this process, 78 papers were identified for a detailed analysis.

2.1.4.4 Classification framework

Five research questions were proposed to be answered by the representative body of literature, 78 papers in our case. To do so, a set of coding parameters and categories were created and based on these, the analysis took place. The next sections contain a detailed description of this categorisation used in our detailed analysis of the 78 papers.

2.1.4.4.1 Industry classifications

For the industry classification, inspiration was drawn from the Global Industry Classification Standard (GICS) (S&P Global & MSCI, 2018) with the addition of two domains, i.e. those of manufacturing and emergency services, as seen in Table 1.

Table 1 Industry classifications

Domain	**Definition**
Aerospace & Defence	Producers of civil or military aerospace and defence equipment, parts, services or products.
Construction & Engineering	Companies engaged in primarily non-residential construction. Includes civil engineering companies and large-scale contractors.
Education Services	Companies providing education services, either online or through conventional teaching methods. Includes universities, correspondence teaching, providers of educational seminars, educational materials and technical education.
Healthcare (Providers, Services & Technology)	Providers of patient healthcare services and companies providing information technology services primarily to healthcare providers.

Manufacturing	Manufacturing companies and related processes like assembly, factory management, maintenance and on- and off-premises training scenarios (health and safety, ergonomics, etc.).
Transportation (Infrastructure & Trucking)	All transportation infrastructure and companies providing primarily goods and passenger land transportation.
Metals & Mining	Companies engaged in the diversified production or extraction of metals and minerals.
Emergency Services	This sector provides a wide range of prevention, preparedness, response and recovery services during both day-to-day operations and incident response. In this paper, the focus is on industrial fire departments and private emergency/medical training service providers.
Utilities (Water & renewables)	Companies that purchase and redistribute water to the end consumer and companies that engage in the generation and distribution of electricity using renewable sources.
Unspecified/generic	The industry is not mentioned, abstracted tasks which hold the potential to be applied to industrial use cases.

2.1.4.4.2 Skills

Industrial skills range from the fine motor skills required of a surgeon suturing an incision to the safety procedures needed when a factory worker operates a piece of machinery. Since existing classification taxonomies around industrial skills are wide-ranging, we propose four categories for this study, as depicted in Table 2. Soft skills were not part of the categorisation, as their useful but generic attributes can be applied in any work setting and thus might dilute our focus on industry.

Table 2 Categories of industrial skills coded

Category	**Definition**
Perceptual motor/ psychomotor skills	Learning skills that require hand-eye coordination to solve the problem.

Procedural skills	Learning of processes and sequences.
Decision-making/ problem-solving	Learning non-linear processes where optimal strategies out of many options need to be selected.
Spatial	Understanding, reasoning and remembering the spatial relations among objects or space.

2.1.4.4.3 Learning theories

Learning theories provide a theoretical framework in which the IVR training and testing occur. A set of learning theories, as they were also exposed from the 78 representative studies, are listed in the Appendix where a table with the theories along with their descriptions is depicted.

2.1.4.4.4 Research design framework

To provide the reader with a comprehensive overview of the methods used to collect the data, we adapted the overview prepared by Radianti et al. (2020), as shown in Table 3. The adaptation is based on the variations of the designs identified in the literature.

Table 3 Research data collection design

Category	**Definition**
Development	A design- or development-oriented study that documents the overall development process.
Experimental design	An experimental/comparative study.
Case study	A study in which data are collected from a bounded system, population or a specific entity.

We also classify the publications according to the type of experimental method they include, as described in Table 4. To aid the reader to reflect on the validity and generalisation of the results, it is important to know the number of participants in the VR experiment. Thus, the average number of participants per condition, which, in the case of between-subjects experimental studies, is the total number of participants divided by the number of experimental conditions, can be seen in the analysis section. Additionally, in the case of within-subjects and preliminary study methods, the number of total participants is considered.

Table 4 Experimental methods

Category	Definition
Between-subjects	Users are assigned to different groups with one or more conditions differing across groups.
Within-subjects	The same group of users is exposed to one or more conditions.
Preliminary study/ survey	There is no experiment, and only survey data is collected, usually for pilot studies.
No study	No user study is performed.

2.1.4.4.5 Data collection methods

IVR training can employ various objective and subjective measures, which are then analysed to understand the effectiveness of the experiment or technology. Objective measures, like task completion time, reaction time or performance scores, form part of our analysis as well, as described in Table 5.

Table 5 Time-based measures

Category	Definition
Reaction time	Time elapsed between a stimulus and a response from the user.
Task completion time	Time elapsed between the start and end of a task.

Regarding the subjective measures, four main categories are included, as seen in Table 6. Examples of techniques that aid researchers in measuring cybersickness are the simulator sickness questionnaire (SSQ) (Kennedy et al., 1993) and the motion sickness scale (Keshavarz & Hecht, 2011). Additionally, for usability measurement, one may use the system usability scale (SUS) (Brooke, 1996) and/or the questionnaire for intuitive use (QUESI) (Naumann & Hurtienne, 2010). For the task load, the most popular choices are the NASA-TLX (task load index) (Hart & Staveland, 1988), the instantaneous self-assessment (ISA) technique (Tattersall & Foord, 1996) and the SIM-TLX (simulation task load index) (Harris et al., 2019). Finally, immersion metrics include the presence questionnaire (Witmer & Singer, 1998) and the Igroup presence questionnaire (IPQ) (Schubert et al., 2001).

Table 6 Data collection methods

Category	**Definition**
Cybersickness	Measures if the user feels nauseated during or after the IVR experience.
Usability	Measures subjective usability data, usually through surveys.
Task load	Measures the cognitive load of the IVR system or learning tasks on the user.
Immersion	Measures the feeling of immersion or presence felt by the user while using the system.

2.1.4.4.6 Data analysis methods

For the analysis of data, a variety of statistical approaches were identified, including ANOVA (analysis of variance), Student's t-test, the Games-Howell post-hoc test, etc. (a detailed overview can be seen in section 2.1.5.3.3). Additionally, when a publication contains no statistical analyses, it is placed in the "no method" category.

2.1.4.4.7 Technologies

We identify the various head-mounted displays (HMDs) used in the publications, including the HTC VIVE, HTC VIVE Pro, Oculus Rift, FOVE HMD, Acer MR and Google Cardboard. In addition to that, we categorise the training space of the IVR system as room-scale (i.e., interactions/training happen in a room-scale space, allowing the user to walk around), arm-scale (interactions/training happen in a limited space within arm's reach) and "not mentioned" in case the interaction/training space is not mentioned or cannot be deduced from the publication. The "unspecified" categorisation is provided when the software or game engine used to create or run the IVR training is not mentioned.

2.1.4.4.8 Biosensors

Biosensors measure bio-signals in the body, including gaze patterns, heart rate, skin conductance, error potentials in the brain, etc. that indicate different aspects of the user's mental and physiological state such as attention, arousal and fatigue. Biosensors available for research are EEG (electroencephalography), GSR (galvanic skin resistance), pupil tracking and HRV (heart rate variability), as seen in Table 7. The papers are also coded for the presence of biosensors as well as their placement on the body (fingers, head, wrists and chest).

Table 7 Categories of biosensors

Category	Definition
Skin conductance	GSR (galvanic skin resistance) indicating user stress levels.
Heart rate signals	Heart rate variability using PPG (photoplethysmography), also known as blood volume pulse sensor.
Brain signals	Electroencephalogram (EEG) electrodes placed on the head (usually the scalp).
Eye tracking	Tracking of pupil size and position usually mounted inside the VR HMD.

2.1.4.4.9 Haptics

Haptics, or specifically computer haptics, refers to the hardware and software enabling the display of haptic stimuli/feedback to the human user (Srinivasan & Basdogan, 1997). For this study, we classify haptics technologies into four categories, namely wearable, pseudo, portable and grounded haptics, as seen in Table 8.

Table 8 Categories of haptic devices

Category	Definition
Wearable	Devices with embedded technology that can be worn on the body to provide force feedback (gloves, exoskeletons) (Pacchierotti et al., 2017).
Pseudo	Objects and surfaces in the training space which require active interaction by the user (e.g., the user grabs a physical cylinder which in the VR environment corresponds to a lever) (Li et al., 2014).
Portable	Non-grounded, mainly user-held devices that have various interfaces for user interaction and whose rotation and position are tracked in 3D space. They may have inbuilt vibration motors for haptic feedback (e.g., Oculus Touch, VIVE Controllers).

Grounded	Mechanical devices which are physically connected to the training space and provide force feedback to the user's hands or fingers (e.g., Geomagic Touch, Novint Falcon) (Biggs & Srinivasan, 2002).

2.1.4.4.10 IVR effectiveness

It is possible to establish the effectiveness of IVR-based training as compared to conventional desktop-based VR or the original physical training scenario. Table 9 shows such a coding scheme alongside the definition of each of the four categories.

Table 9 Measures of IVR training effectiveness

Category	Definition
Effective	IVR-based training is more effective than the alternative (desktop, physical training, etc.).
Not as effective	IVR-based training is less effective compared to the alternatives.
Inconclusive	There is no definite evidence to suggest the better training scenario.
Not applicable	A comparison between IVR and non-IVR training is not described or does not involve any user study at all.

2.1.4.4.11 IVR remote training

To highlight the applicability of IVR applications against remote training, the following categorisation was applied, as seen in Table 10. In the table, the categories and a description for each are provided.

Table 10 Categories of IVR training

Category	Definition
Autonomous Training	Instructions are available inside the VR environment in the form of audio, video and text. The physical presence of the trainer is not required.
Guidance/Monitoring	Continuous involvement of a trainer or operator is needed. This is in the form of continuous physical assistance or feedback to the user, and also in the form of monitoring of the VR training.

Initial Guidance/Setup	Only initial involvement by a trainer or operator is needed to assist the user in using the VR software and hardware environment.
Remote Guidance	An expert or trainer is present at a remote location and trains the user in the form of an avatar and/or with the help of audio, visual or text cues.
Remote Peer Collaboration	Learning is facilitated between a group of remotely located users (trainees) connected via avatars and/or audio, video or text communication.

Additionally, a categorisation of the technical requirements for each application is exposed with the following three categories, as depicted in Table 11. In the table, the three categories indicate the required devices and define the parameters of the devices' use.

Table 11 IVR training applications' technical requirements

Category	**Definition**
VR Headset Kit	Standard VR kit including HMD (Head Mounted Display), integrated or external trackers, handheld controllers, PC and peripherals, and/or eye tracking attachments.
VR Headset Kit + 1 Additional Device	VR headset kit along with at least one additional device (biosensor, grounded haptic device, pseudo-haptic object, portable haptic device, wearable haptic device, handheld game controller)
VR Headset Kit + Multiple Devices	VR headset kit along with at least multiple devices (biosensor, grounded haptic device, pseudo-haptic object, portable haptic device, wearable haptic device, handheld game controller)

2.1.5 Results

2.1.5.1 Trends in skills training *(RQ1: What types of skills does immersive VR training provide in the industry?)*

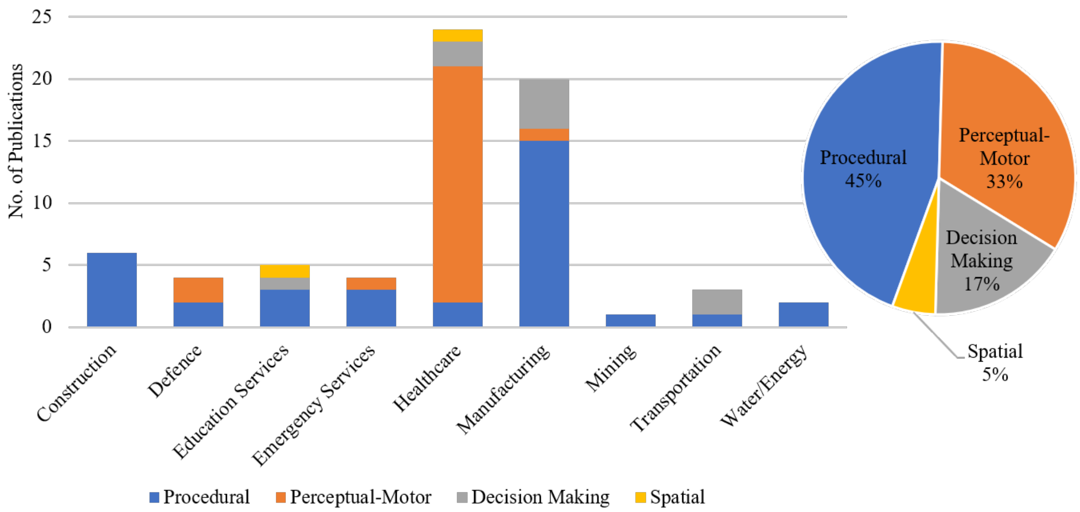


Figure 16 Distribution of skill types across industries and publications

Procedural skills were found to be the most dominant skill type trained at 45% (N=35), followed by perceptual-motor skills at 33% (N=26), decision-making skills at 17% (N=13) and, finally, spatial skills at 5% (N=4), as depicted in Figure 16. In the subsequent sections, a detailed analysis will be presented.

Procedural skills

Of the 35 publications focusing on procedural skills, half of the publications (N=12) were from the manufacturing sector. Some illustrative examples include training factory workers to learn the steps for working with industrial robots to lay tapes for building aerospace composite parts Matsas and Vosniakos (2017), Wang et al. (2019) building an IVR training system that enabled a remote expert in VR to train a local worker in a task involving the assembly of a vice in a manufacturing setting, and Winther et al. (2020) providing a sequential pump maintenance task to novice apprentices in an industrial setting.

Perceptual motor skills

Among all publications focusing on perceptual-motor skills (N=26), most were in the healthcare domain (N=19), focused on teaching some variety of fine or gross motor skills. Popular examples of this training method include four publications on variations of endoscopic surgical skills (Frederiksen et al., 2020; Huber et al., 2017; Jain et al., 2020; Li et al., 2020) and three on hip replacement surgery (Hooper, Tsiridis, Feng, Schwarzkopf, Waren, Long, Poultsides, Macaulay, Papagiannakis, Kenanidis, et al., 2019; Knopp et al., 2018; Panariello et al., 2019).

In addition to healthcare-related publications, our findings on publications on perceptual-motor skills also include two papers in the defence domain, one in emergency services and manufacturing and three papers in the unspecified/generic domain. In the defence domain, the two publications addressed shooting training (Muñoz et al., 2019) and an abstract scenario (Kohli, 2010) where differently shaped virtual objects were mapped to a single physical surface by subtly changing the location of the user's fingers. The publication in the emergency domain considered training in the operation of a fire hose using a haptics-enabled simulator (Nahavandi et al., 2019), while the publication in the manufacturing domain involved a bimanual burr puzzle assembly task (Murcia-Lopez & Steed, 2018). In the unspecified/generic domain, S. Xiao et al. (2020) explored a drawing task, Harris et al. (2019) had a virtual block stacking task and Škola et al. (2019) had participants shoot down asteroids using brain motor imagery detected through EEG sensors.

Decision-making skills

13 publications focus on decision-making skills, half of them clustering in the manufacturing and unspecified/generic domains with four publications for each domain. Next are healthcare and transportation with two and finally education and emergency services with one publication each. A representative publication from the manufacturing sector is a study describing scenarios for training participants in identifying flaws in the layout plan for assembling a ceiling-mounted installation system (Hirt et al., 2019), and another is a study that had the training system learn complex assembly procedures from experts and then train novices to solve the assembly task using visual cues (Roldán et al., 2019).

A healthcare example is an IVR system for training medical students in deciding between surgical treatment plans based on radiograph scans (Sakowitz et al., 2019), and in transportation, an IVR-based flight simulator trained participants in managing resources like fuel levels in different tanks (Luong et al., 2020). The only publication in education services focusing on decision-making skills involved an IVR-based visualisation of students' eye gaze to train teachers in identifying distracted students (Rahman et al., 2020), and the only example in the emergency services domain involved participants making route navigation decisions on exiting a virtual museum after a fire breakout (Cao et al., 2019).

Spatial skills

Only four publications refer to spatial skills. In more detail, Pollard et al. (2020) used a VR-based scavenger hunt scenario to investigate the effects of levels of immersion on object recognition and discrimination tasks. Sun et al. (2019) measured the effect of the learning environment on spatial skills among high- and low-spatial-ability users by comparing an IVR environment to a slide presentation scenario, whereas Srivastava et al. (2019) focused on spatial learning for navigating a virtual environment across IVR and desktop-VR conditions, and Zinchenko et al. (2020) utilised IVR to train participants to identify anatomical features of a virtual heart.

2.1.5.2 Learning Theories (*RQ2: Which learning theories are used in IVR industrial skills training?*)

In 14% (N=11) of the total number of publications, the authors explicitly mentioned 13 learning theories, including, for example, constructivist learning theory (Akanmu et al., 2020), cognitive load theory (Frederiksen et al., 2020; Sun et al., 2019) and cognitive theory of multimedia learning (Meyer et al., 2019). The theories used in the papers are listed below (definitions are available in the Appendix).

- *Constructivist Learning Theory:* Akanmu et al. use Constructivist learning theory to design their IVR- based posture training system based on the learner's current and previous experience (Akanmu et al., 2020).
- *Cognitive Load Theory*: Sun et al. applied Cognitive Load Theory (CLT) to measure the effectiveness of a spatial ability training system on users with high and low spatial ability (Sun et al., 2019), while Frederiksen et al. used CLT to design an experiment comparing the cognitive load between IVR and non-immersive conditions (Frederiksen et al., 2020).
- *Cognitive Theory of Multimedia Learning:* Meyer et al. (2019) use the cognitive theory of multimedia learning to design their training scenarios where the content is designed to reduce intrinsic load (using the pre-training principle) while the experiment measures the impact of extrinsic cognitive load arising from the method of instruction (IVR vs desktop).

- *Deliberate Practice Theory (DPT)*: Butt et al. (2018) designed a game based on IVR for training urinary catheterization which used principles of DPT and compared learner performance to a non-game, non-IVR group that learned the same skills in the traditional manner.
- *Fitts & Posner Stages of Motor Skill Acquisition*: Carlson et al. use this framework to design and analyse implications of the learning in their scenario involving users learning how to solve an assembly puzzle across physical and IVR conditions (Carlson et al., 2015). Winther et al. (2020) also mention this theory as a background in their development of an IVR pump maintenance trainer.
- *Fowler's Theory of Experiential Learning:* Simeone et al. (2019) use this theory to create an IVR training system to train users in the underlying concepts of IVR design.
- *Experiential Learning:* Akanmu et al. (2020) used Experiential Learning theory to motivate their study participants to take ownership of their skills through hands-on learning activities.
- *Gagne's Flow model:* Wu et al. (2020) used Gagne's flow model with its 9 instructional events for increasing learning performance among nurses (less needlestick injury).
- *Insight Learning:* Collins et al. (2019) designed their bio-sensor-based system to detect insights that are correlated with learning according to Insight Learning Theory.
- *Knowles Theory of Adult Learning:* Akanmu et al. (2020) adapted their IVR training system to focus on adult learners who need reasons for learning, and who learn better when they perceive the skills to be relevant to their daily lives.
- *Lander's theory of gamified learning:* Pollard et al. (2020) cite Landers' model in the design of their IVR system, where immersive technology does not replace instructional content but helps in influencing the behaviour of the learner and influence learning impact.
- *Pre-training principle:* Meyer et al. lessen the external cognitive load (Mayer & Moreno, 1998) of IVR training by using the pre-training where the names, shape and colour of cellular components were introduced before the actual training (Meyer et al., 2019).
- *Thorndike transfer of practice:* Winther et al. (2020) mention Thorndike transfer of practice and its relevance to virtual training environments which can minimize contextual change and lead to better learning recall.

2.1.5.3 Research design (*RQ3: Which research designs, data collection methods and data analysis methods are utilised in IVR industrial skills training?*)

The data revealed that 82% of the publications applied an experimental design approach (N=64), including user surveys and between-/within-subjects studies, followed by 18% of the publications focusing on the design and development of IVR training without emphasising validation (N=14). Only one instance of a case study approach was reported in the data (Shamsuzzoha et al., 2019). Furthermore, the between-subjects validation method had the highest representation at 62% (N=48), followed by the within-subjects and preliminary study/survey with a share of 15% (N=12) each. Six publications did not mention any kind of validation (no study). In papers using between-subjects validation, healthcare had the highest representation with 20% (N=16), followed by manufacturing at 17% (N=13).

2.1.5.3.1 Data collection

Cybersickness

18% (N=14) of the papers measured cybersickness. The measures used were the simulator sickness questionnaire (SSQ) (N=8) (Kennedy et al., 1993) and the motion sickness scale (N=2) (Keshavarz & Hecht, 2011). The other four papers collected non-standardised measurements of sickness that did not use pre-validated measures of cybersickness, but included questions related to it as part of the general user questionnaire employed in the studies. For example, Pérez et al. (2019) included sickness as one of the 12 questions in their user validation survey.

Usability

In the body of literature examined, 20% (N=16) of the papers measured usability. Among these papers, the system usability scale (SUS) was the most widely used (Brooke, 1996) with five papers applying it. The questionnaire for intuitive use (QUESI) was used by one publication (Li et al., 2020). The remaining papers measured usability using non-standardised methods; for example, Pérez et al. (2019) included usability questions as part of their survey, which also included other measures like presence.

Task load

17% (N=13) of the papers measured task load/difficulty (defined in Table 6). Among these papers, the NASA-TLX (task load index) was the most commonly used, represented in eight of the publications measuring task load measures (N=13). Interestingly, in addition to the well-established NASA-TLX, Harris et al. (2019) introduced the SIM-TLX measure, an adaptation of the NASA-TLX which has been validated for measuring task load in VR simulators. One publication used an onscreen adaptation of the instantaneous self-assessment (ISA) method to measure task load during the IVR training (Tattersall & Foord, 1996). Another publication used an auditory stimulus (beep sound) and measured the reaction time between the beep and the user's response to the stimulus to measure cognitive load (Frederiksen et al., 2020).

Immersion

18% (N=14) of the papers measured aspects of immersion, presence, or embodiment. We observed a fusion of standardised measures (i.e., validated in previous literature), adapted measures (from a previously validated measure) and non-standardised measures (non-validated questions, usually asked with other data collection measures).

- *Standardised measures*. Three papers referred to Witmer and Singer's presence questionnaire (Witmer & Singer, 1998), and three followed the Igroup presence questionnaire (IPQ) (Schubert et al., 2001), Nichols et al.'s measurements of presence (Nichols et al., 2000) and an adapted form of the immersive tendencies questionnaire (ITQ) (Witmer & Singer, 1998), one for each paper.
- *Adapted measure.* Škola et al. (2019) created an embodiment questionnaire based on Botvinick and Cohen's rubber hand illusion questionnaire (Botvinick & Cohen, 1998) and Longo et al. (2008)'s psychometric questionnaire for embodiment.
- *Non-standardised measures*. Four papers used non-standard measures of immersion particular to their use cases. For example, Simeone et al. (2019) asked participants to rate their sense of presence while using the IVR system (for teaching IVR theoretical concepts) along with questions about preference and self-assessment of performance.

Time-based measures

- *Task completion time.* 42% (N=33) of the papers measured task completion time. Fourteen of these papers are from the manufacturing sector, followed by healthcare with nine, construction and water/energy sectors with two each and emergency services, education services and transportation with one each (see Figure 17). Three publications using time-based measures were identified in the unspecified/generic domain.

- *Reaction time.* Five papers used the reaction time metric, with representations in healthcare (N=2), the unspecified/generic domain (N=2) and in educational services (N=1) (see Figure 17).

Score-based measures

- Slightly more than half of the papers (N=41) used some variety of scoring metrics, including keeping track of the number of errors (N=11), successful steps completed or accuracy (N=20) and other measures like performance in knowledge retention tests (N=10). The healthcare domain used scoring metrics the most with fourteen publications, followed by manufacturing with nine, education services with four, construction with three, and two each for the transportation and water/energy domains followed by one each for the other domains. Finally, four publications using score measures were identified in the unspecified/generic domain (see Figure 17).

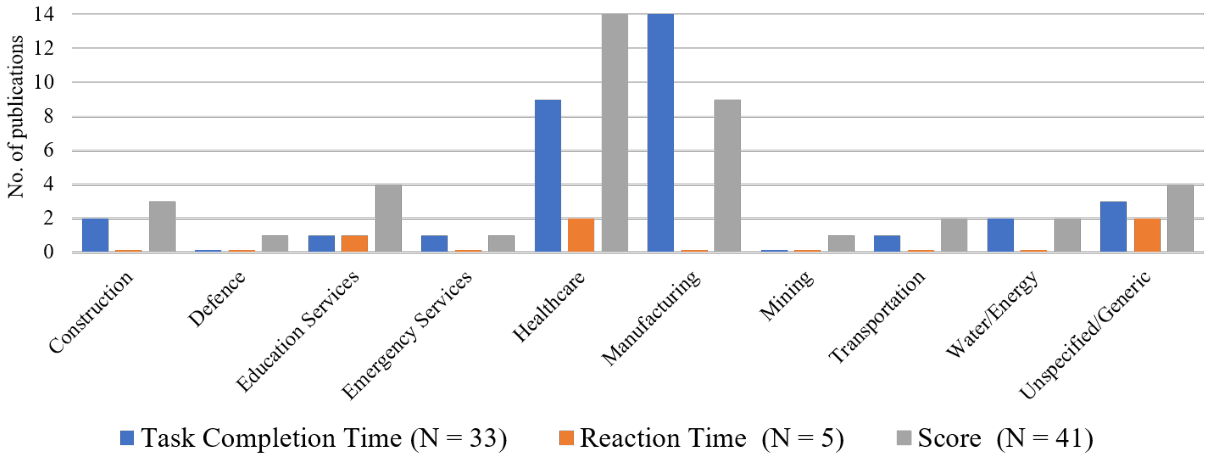


Figure 17 Use of task completion time, reaction time and scoring metrics across industries. Note: each publication may be included in more than one category

2.1.5.3.2 Experimental methods and study size

The average number of study participants per experimental condition in our data set is 17.55 participants. However, this may not be as informative to the reader as the average number of participants per condition per industry. Applying this perspective, the emergency services sector takes the lead with an average of 75.75 participants per condition, followed by the education services and water/energy sectors with 22.9 and 22.75 respectively. At the bottom of Figure 18, the total number of publications per industry appears in circles. This was added to put the average study size in perspective. Thus, for example, it is evident that emergency services have the highest average number of participants (63.8), yet there are only five publications in this domain. On the other hand, healthcare has on average around 16 participants per condition although it has the largest number of publications in the database.

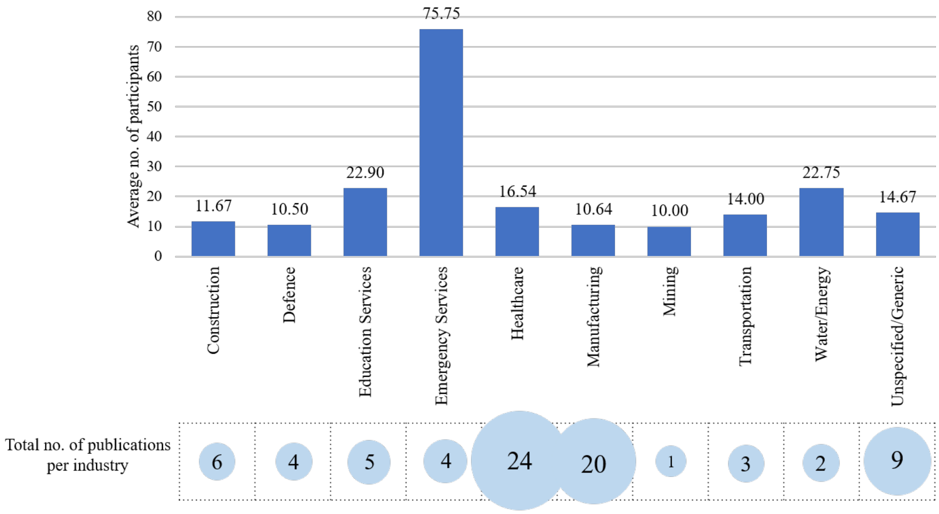


Figure 18 Average number of participants per condition per industry

2.1.5.3.3 Data analysis

As for the analytical method applied, we see that 82% of the publications follow an experimental design approach (N=64), 17% (N=13) are development-oriented and only one paper takes the form of a case study (Shamsuzzoha et al., 2019). Figure 19 shows the data on the average participant size, the experimental approach, the research method, and the data analysis methods. On the left side of the figure, the average number of participants per category is divided into three distinct categories, based on the data discussed in the previous section. The first group, '1-20' denotes studies with 1 to 20 participants per condition, with a similar logic for the '25-50' and '51-200' groups. The figure also shows a) the data from the experimental approach used in the publications (as described in Table 4), b) the research data collection method (described in Table 3) and c) the data analysis method used (as referred to in section 2.1.4.4.6). The circled numbers inside the grids in the bottom half part of Figure 19 denote the numbers of publications where these factors overlap. For example, the biggest number bubble in the figure is 48 (bottom left), which denotes that 48 publications follow an experimental design and are mapped vertically with the between-subjects category. If there are no publications, the grid is empty. For example, there are no publications that use the development research method and also have an average participant size in the range of 51-200.

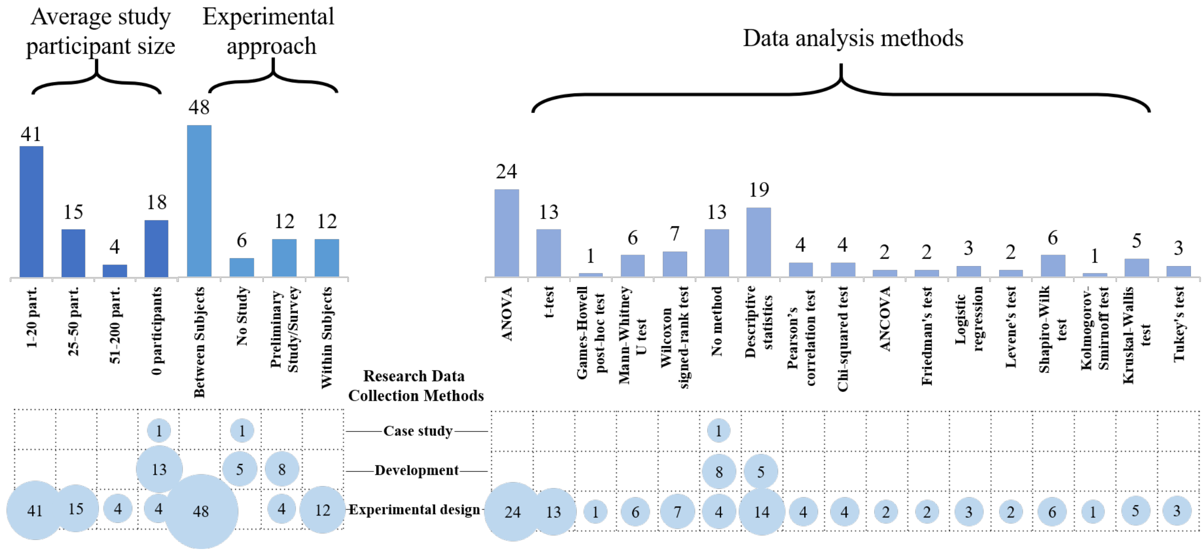


Figure 19 Comparison of research methods, study size, experimental approach, and data analysis methods. Note: each publication may be included in more than one category

Figure 19 reveals some interesting trends. For example, the experimental approach most commonly used in experimental design is between-subjects (N=48), while about 41 publications in experimental design have an average participant size of less than 20. As far as statistical measures are concerned, we see that ANOVA is the most widely used measure with 31% (N=24), which is followed by the use of descriptive statistics at 26% (N=20). Additionally, 13 publications use no statistical measure at all. The only statistical measure used by development-oriented publications is descriptive statistics (N=5).

2.1.5.4 VR, biosensors, and haptic technologies *(RQ4: What HMD technologies, biosensors and haptics are utilised in immersive VR industrial skills training?)*

2.1.5.4.1 VR technologies

Use of head-mounted displays

On the subject of the display technologies utilised in IVR training in the industry, Figure 20 depicts on the right a pie chart with the distribution of 10 HMD brands. The most widely used HMD was the HTC VIVE with 45% of the papers (N=35) reporting on its use, followed by the Oculus Rift used by 26% of the papers (N=20). Eight papers use the slightly more advanced version of the HTC VIVE – the VIVE Pro. Seven papers do not mention HMD use at all.

In Figure 20, one may observe, in the healthcare domain, the Oculus Rift (N=10) and the HTC VIVE (N=9) dominate the sector. Likewise, in the manufacturing domain, the HTC VIVE dominates with twelve publications employing its features, followed by the HTC VIVE Pro with three papers.

Training space

In 62% of the publications (N=48), the VR training space was limited to arm's scale, and the user was either seated or standing (without moving around). Another 36% used room-scale VR (N=28), where different tracking technologies enabled tracked navigation around a small area. In more detail, tracking was enabled by the use of cameras placed around the training space tracking the position and orientation of the HMD and the handheld controllers, if there were any, and this was evident for all publications using HMDs from HTC and Oculus. Exceptions in user tracking exist, like Matsas and Vosniakos (2017), where a Microsoft Kinect depth-sensing camera coupled with an eMagin Z800 3Dvisor HMD was used for user tracking.

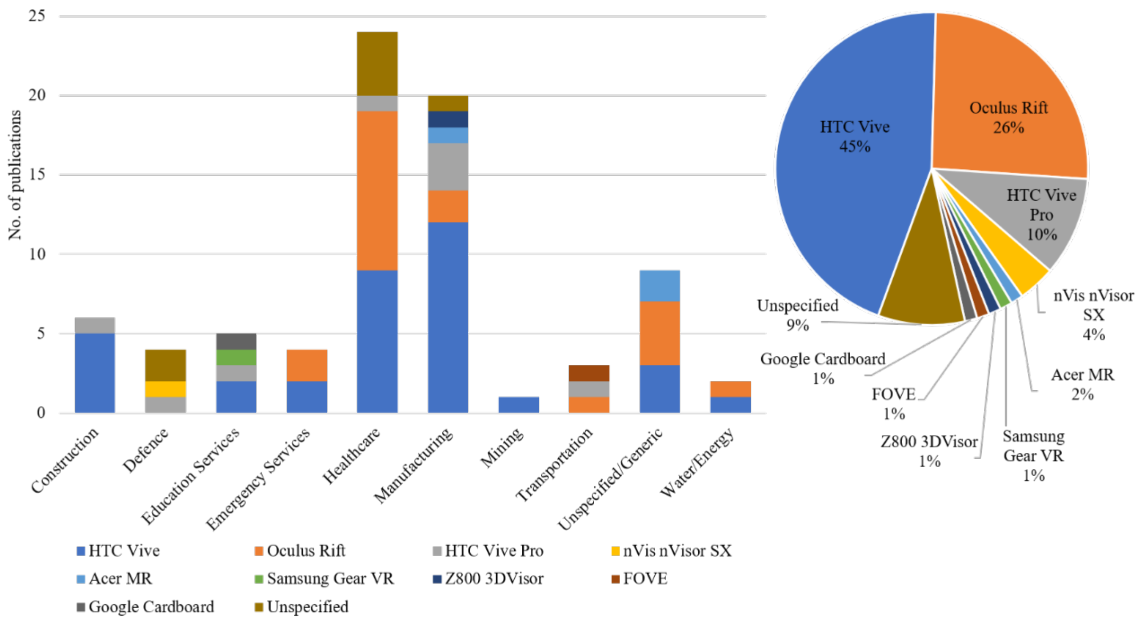


Figure 20 Distribution of HMDs across industries

2.1.5.4.2 Use of biosensors

In the reviewed body of literature, 18% (N=14) of the publications used biosensors measuring brain signals (N=8), heart rate (N=3), eye tracking (N=6), temperature (N=1) and skin conductance (N=1). Three publications used the biosensor data for adapting the training to the user. As shown in Figure 21, the unspecified/generic domain holds the most publications (N=3), followed by manufacturing, education services, emergency services with two, and the construction, healthcare, defence, transportation, and water/energy industries holding one publication each. It is notable that all 14 publications using biosensors are quite recent (from 2018 and onwards), and none of them combines biosensors with any form of haptic feedback.

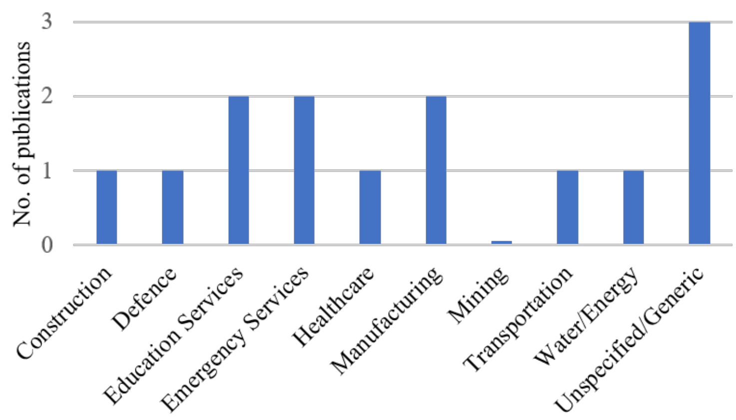


Figure 21 Use of biosensors across industries

Collins et al. (2019) used skin conductance, temperature, pulse and heart rate data to understand the emotional responses felt by users when they gained insights into the solution while performing an abstract task of manipulating a hypercube, whereas Škola et al. (2019) created an EEG-based adaptive training system for increasing the user affect in a gamified virtual environment to provide motor imagery-related biofeedback. This increase in user affect was used to train the user to control a spacecraft in an asteroid shooting game. Sun et al. (2019) used EEG signals to identify the effect of the presentation medium (IVR vs presentation slides) on cognitive load. Baceviciute et al. (2020) used EEG measures in addition to other evaluation techniques to measure the comparative effectiveness of different representations of text and audio in an IVR learning environment.

Biosensing technologies and locations on the body

Figure 22 shows the body locations of all the biosensors mentioned in our database. All cases with EEG use had the sensors placed on the head, and the cases with gaze tracking also had the eye tracking hardware mounted on the head as part of the VR headset itself. When it came to ECG sensors, Muñoz et al. (2019) measured HRV (heart rate variability) by placing the Polar H10 sensor strap on the chest, while Collins et al. (2019) used the Empatica E4 wristband sensor to measure the same along with electrodermal activity, skin temperature and blood volume pulse. Longo et al. (2019) did not mention which heart rate variability sensor was used. The gaze tracking features of the HTC VIVE Pro Eye was used for identifying distracted learners (Rahman et al., 2020) and for foveated rendering in procedural task training

(Radkowski & Raul, 2019), while the gaze tracking features of the FOVE HMD was used in one case (Lang et al., 2018). Another case used eye tracking for IVR-based welding training, but the model of eye tracking hardware used was not mentioned (Torres-Guerrero et al., 2019). In our database, various EEG sensors (N=7) were used for sensing brain signals: Emotive EPIC for measuring user stress and concentration (Torres-Guerrero et al., 2019), Brain Products ERP recorder and analyser (Sun et al., 2019), Neuroelectrics Enobio EEG system (Škola et al., 2019), Liveamp EEG cap (Dey et al., 2019), Muse BCI EEG system (Muñoz et al., 2019) and the Advanced Brain Monitoring (ABM) X-10 (Baceviciute et al., 2020). The other brain-signal sensing technology in the database is in Shi, Zhu, et al. (2020), where a functional near-infrared spectroscopy (fNIRS) device (a NIRSportTM worn on the head) is used along with an HMD integrated eye tracker (Tobii) to measure stress levels among users who were being trained on industrial shutdown maintenance.

Biosensors and skills

Out of the 14 papers using biosensors, four papers had decision-making (Collins et al., 2019; Dey et al., 2019; Lang et al., 2018; Rahman et al., 2020)) and seven had procedural skills as the dominant skill taught (Baceviciute et al., 2020; Longo et al., 2019; Shi, Du, et al., 2020; Shi, Zhu, et al., 2020; Torres-Guerrero et al., 2019; Zou et al., 2016) followed by perceptual-motor skills with two papers (Muñoz et al., 2019; Škola et al., 2019). Only one publication focused on spatial skills (Sun et al., 2019).

Biosensors and training space

Of these 14 studies, ten took place in an arm-scale training space. Among the papers using a room-scale training space, Collins et al. (2019) had a room-scale configuration and measured heart rate, skin conductance and temperature of users being trained in a hypercube manipulation task; however, it is not clear from the paper, if the users moved around the environment. In the second publication using biosensors in a room-scale space, Longo et al. (2019) used a heart rate sensor to measure stress levels in trainees undergoing emergency preparedness training.

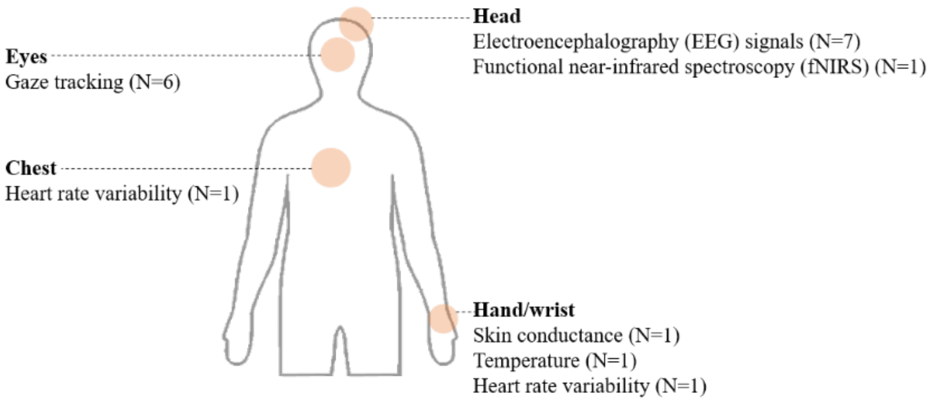


Figure 22 Distribution of biosensors across the body as mentioned in our database

2.1.5.4.3 Haptic feedback

35% of the publications (N=27) used a variation of haptic feedback. Of these papers utilising haptics, more than a third (N=10) used grounded haptics, eight used portable haptics, seven used pseudo-haptics and two used wearable haptics, as appears from Figure 23. Almost two-thirds of these papers (N=16) belonged to the healthcare domain, followed by the manufacturing sector with three and the defence sector with two, while the remaining industries each hold a paper.

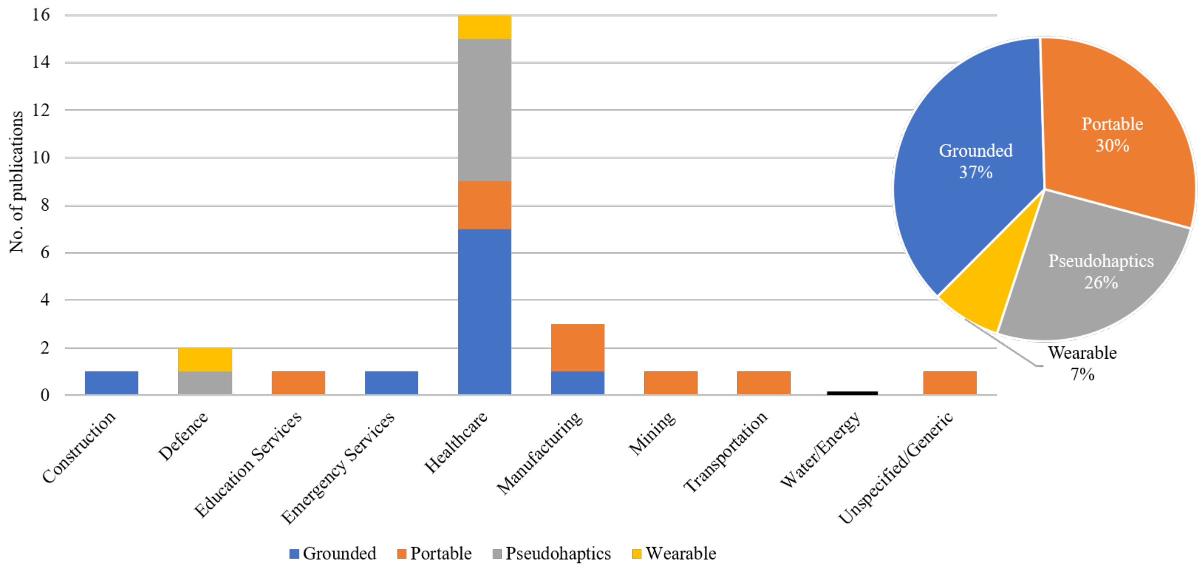


Figure 23 Use of haptics across publications and industries

Grounded haptics

Of the 10 publications using grounded haptics, seven papers were from the healthcare domain and one each from the construction, emergency services and manufacturing domains. Half of them used off-the-shelf haptic hardware. Three employed the 3D Systems Geomagic Touch (Carlson et al., 2015; Vaughan et al., 2019; Xin et al., 2020), one used the Geomagic Touch X (X. Xiao et al., 2020) and one used the LapMentor III surgical platform (Li et al., 2020). One publication customised a KUKA industrial robot with haptic feedback for surgical training (Pelliccia et al., 2020), while three built novel haptic devices (Bin et al., 2019; Durai et al., 2019; Nahavandi et al., 2019). Lastly, one publication adapted an existing haptic device but did not describe it (Sainsbury et al., 2020).

Portable haptics

Seven publications in our database used portable haptic devices, represented by the HTC VIVE Controller (N=5) and the Oculus Touch controller (N=2). Applications included simulating tool interaction forces for training in pump maintenance (Winther et al., 2020), utilising the VIVE Controller's vibration for training in rock scaling operations (Liang et al., 2019), and providing performance-related feedback for motor imagery training using vibrations from the Oculus Rift Touch controller (Škola et al., 2019).

Wearable haptics

Two cases, one in the defence (Duggan et al., 2019) and one in the healthcare (Butt et al., 2018) domain, used wearable haptics. These included glove-based haptic systems but did not name the devices.

Pseudo-haptics

Seven publications employed pseudo-haptics, with two using physical representations of the human body that corresponded to their virtual counterparts in IVR (Almousa et al., 2019; Jain et al., 2020), three used the minimal physical forces arising from the use of a Simball joystick device for surgical simulation (Frederiksen et al., 2020; Huber et al., 2017; Huber et al., 2018) and another mapped virtual objects in IVR to physical objects in a military training context (Kohli, 2010). One publication applied the term pseudo-haptics outside our defined context, i.e. a virtual hand that did not penetrate a virtual body (no physical interaction was involved) (Bálint et al., 2019).

Haptics and VR training space

Of the 27 papers using haptics, 22 were in an arm-scale space, as shown in Figure 24. Grounded haptics comprises almost half of all arm-scale training cases (N=9) in the database. In room-scale training spaces, three cases of portable haptics (Liang et al., 2019; Schwarz et al., 2020; Winther et al., 2020) can be observed as well as one case each in pseudo-haptics (Jain et al., 2020) and grounded haptics (Nahavandi et al., 2019).

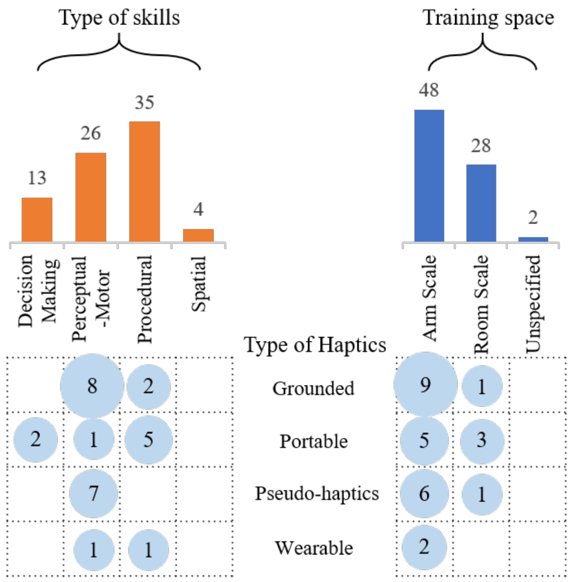


Figure 24 Types of haptics used in relation to training space and skills taught

2.1.5.5 Effectiveness of IVR training *(RQ5: What levels of effectiveness of IVR are reported in industrial skills training?)*

More than half of the publications in our database (N=40) conveyed the extent of the effectiveness of IVR compared to another VR modality (CAVE, desktop VR, physical training, etc.) and reported the result as either effective, not as effective, or inconclusive. The pie chart in Figure 25 shows the overall distribution of the effectiveness of IVR training systems among all publications. In parallel, the bar chart shows the distribution of IVR effectiveness outcomes across industry sectors (without the not-applicable publications).

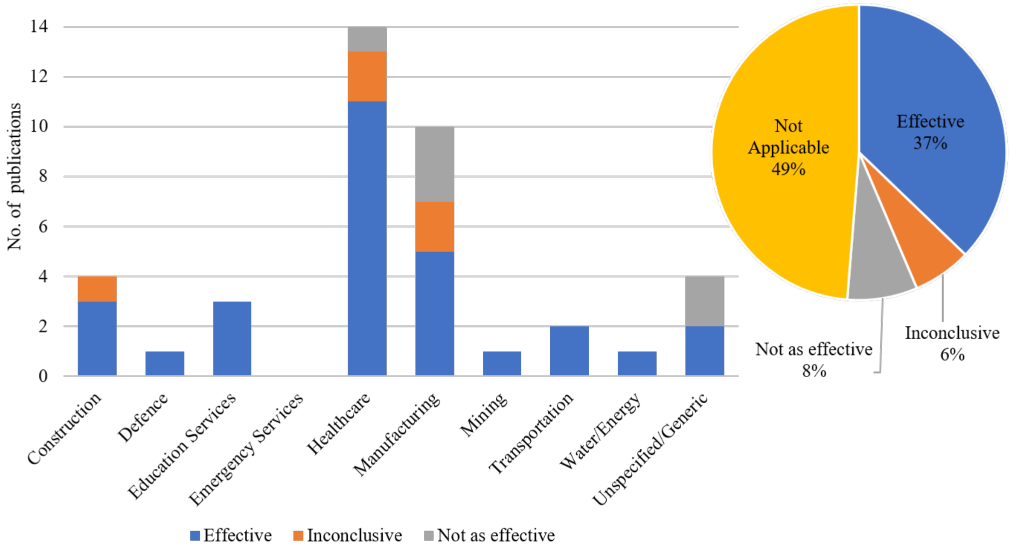


Figure 25 Number of publications across industries with respect to IVR effectiveness. The industry distribution does not include the "not applicable" domain.

- ***Effective***. Of the publications that examined the effectiveness of IVR training, the majority (N=29) concluded that IVR training was effective. For example, Buttussi and Chittaro (2018) compared learner performance in a task under three conditions: high-fidelity IVR, medium-fidelity IVR and desktop VR. Higher fidelity was shown to increase both engagement and presence. Sun et al. (2019) used IVR and presentation slides on a desktop to teach astronomy concepts to both high- and low-spatial-ability users. Interestingly, they reported that IVR benefitted low-spatial-ability users by reducing their cognitive load.
- ***Not as effective***. Six publications found that IVR training compared to non-IVR conditions. For example, Frederiksen et al. (2020) found that in a laparoscopy trainer, IVR-based training resulted in poorer performance and more cognitive load than desktop VR. Similarly, in an assembly task, Barkokebas et al. (2019) found that the control group using an instructional manual took less time and committed fewer errors than the IVR group.
- ***Inconclusive***. Five papers that compared IVR training to other modalities found the results to be inconclusive.

The rest of the publications, comprising 49% of the database (N=38), consisted of IVR comparison studies between different IVR modalities, preliminary studies/surveys (with no results) and studies that did not conduct any experiments. For example, Buń et al. (2019) compared IVR across both high- and low-fidelity settings without any non-IVR conditions.

Design attributes and effectiveness

One may find value by observing the design features, i.e., the distinctive attributes that compose the IVR training, in publications that explicitly reflect on effectiveness, as seen in Figure 26. That is, papers that discuss or reflect on the effect that made their IVR training systems (N=33) a) more effective than non-IVR systems (positive), b) as effective to non-IVR in performance (inconclusive) or c) inferior to non-IVR training (negative). By no means is this a direct recipe for effectiveness, but rather a reflection point based on a frequency count of the representative publications identified. This is more evident, as the same attribute, for example, immersive features, can be seen to be mapped with both positive and negative effects.

- ***Object interaction realism.*** We broadly define *object interaction realism* as the fidelity of the interaction with digital objects in the virtual environment.
  - ***Positive.*** Two cases discussed how the fidelity of haptic interaction affects IVR performance. Xiao et al. (X. Xiao et al., 2020) described realistic haptic interaction in their IVR condition as beneficial, while Carlson et al. (Carlson et al., 2015) found that participants in their IVR condition benefited from the slightly lower level of haptic fidelity as compared to physical training.

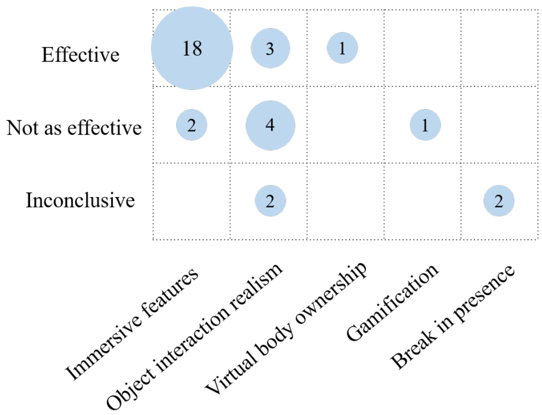


Figure 26 Mapping of design features to IVR effectiveness

  - ***Negative.*** Four cases described their IVR condition to be not as effective as physical training and listed the lack of realistic haptic interaction and other cues as reasons (Barkokebas et al., 2019; Frederiksen et al., 2020; Koumaditis et al., 2020b; Winther et al., 2020).
  - ***Inconclusive.*** Murcia-Lopez and Steed (2018) and Schwarz et al. (2020) reported IVR and physical training conditions to be equally effective, with Murcia-Lopez

and Steed finding the participants in the IVR training condition to spend more time making physically plausible interactions even though there was no haptic feedback. Schwarz et al. hypothesize that the lack of directional haptic feedback affected training in the IVR condition.

- ***Immersive features.*** Immersive VR has certain features that make it distinct from desktop VR, including the angle of view, three-dimensional interaction, head and body tracking and immersive stimuli.
  - ***Positive.*** 19 publications mentioned immersive features to be among the factors benefitting IVR training over other conditions.
  - ***Negative.*** Srivastava et al. found that the increased cognitive load during the IVR training made the training less effective when compared to other modalities (Frederiksen et al., 2020; Meyer et al., 2019; Srivastava et al., 2019), while Ragan et al. concluded that the inability to see one's own body in IVR may have reduced IVR effectiveness (Ragan et al., 2017).

- ***Break in presence.*** This refers to any factor that can make the user lose their perceived feeling of presence in the virtual environment, for example when they are asked to remove the HMD in between the IVR experience to fill in questionnaires (Putze et al., 2020).
  - ***Negative.*** One publication reported that the paper-based assessment in between the IVR training broke immersion (Sakowitz et al., 2019), while another stated the ergonomics of their HMD as a reason (Huber et al., 2018).

- ***Virtual body ownership.*** The acceptance of and identification with the virtual avatar are called the illusion of "virtual body ownership" (Waltemate et al., 2018).
  - ***Positive.*** Škola et al. (2019) reinforced the effect of virtual embodiment in their motor imagery training by using a realistic human-like avatar from the first-person perspective to induce better illusions, which in turn affected learning performance.

- ***Gamification.*** Gamification refers to features usually found in games like displays of score or elements of feedback solely intended to increase user motivation.

- ***Positive.*** Only one publication indicated that gamified features in their IVR training led to better engagement (Butt et al., 2018).

Length of studies

Of the publications measuring the effectiveness of IVR (N=40), we found that all performed immediate testing and that only nine papers measured the effectiveness of the same IVR training over long(er) periods. These nine had a between-subjects methodology (described in Table 4). These publications measured the change in knowledge retained after one week (Liang et al., 2019; Meyer et al., 2019) and two weeks (Blumstein et al., 2020; Butt et al., 2018; Buttussi & Chittaro, 2018; Carlson et al., 2015; Lang et al., 2018; Murcia-Lopez & Steed, 2018; Sakowitz et al., 2019), while one publication also measured perceived levels of enjoyment (Meyer et al., 2019). All of them found that the relative advantage of any modality (IVR, desktop, physical) over another was preserved over periods, except for two cases. More specifically, in Carlson et al. (2015), participants in a physical training condition initially outperformed those in IVR in terms of knowledge retention, but after two weeks, this effect was reversed, and in Meyer et al. (2019), participants in a desktop VR condition initially reported more enjoyment than those in the IVR condition, but after a week, this effect was reversed as well.

2.1.5.6 IVR training remote applicability *(RQ6: How applicable are current immersive VR applications to be provided as remote training solutions?)*

As depicted in Figure 27, most of the IVR applications (N=60) identified in our data can be classified as having autonomous IVR training (*Autonomous Training*), followed by eight publications (N=8) that report the presence of a trainer or operator during the training process to either guide or monitor the user (*Guidance/Monitoring*). Six publications (N=6) require a trainer or operator to set up the hardware and software (*Initial Guidance/Setup*) and four publications (N=4) mention remote training with three requiring a remote trainer (*Remote Guidance*) and one requiring peer learners located remotely (*Remote Peer Collaboration*).

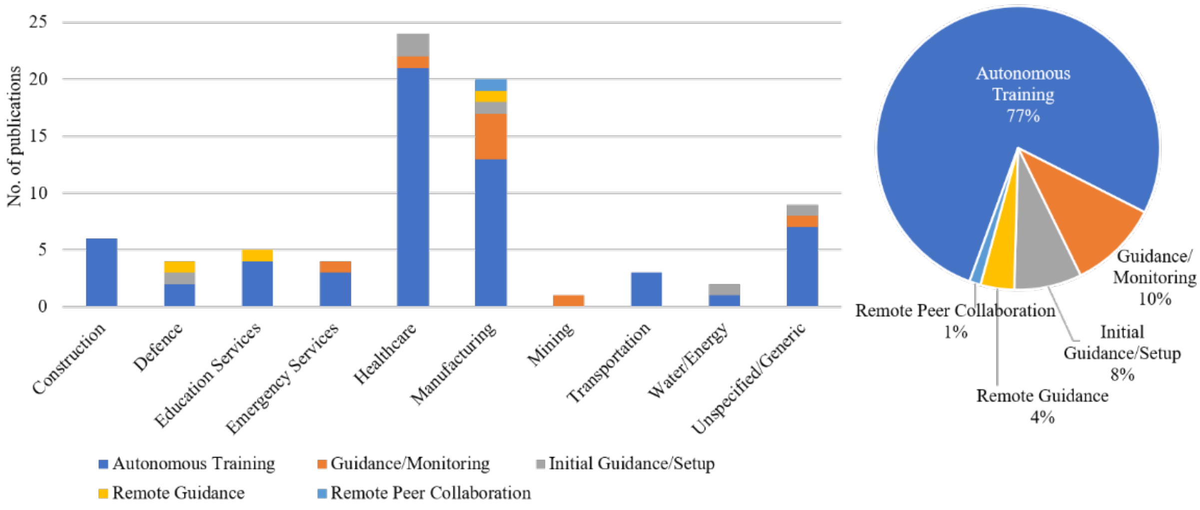


Figure 27 Trends in remote training in IVR across publications and industries

One may observe that the majority of the industrial domains, with one exception (i.e., the mining domain), provide half or more than half of their applications as autonomous training. For example, in the healthcare domain, 21 applications (N=24 total) can be characterised as autonomous training solutions and in the manufacturing domain the majority, i.e., 13 applications (N=20 total), are autonomous IVR training solutions.

The majority of IVR applications requiring continuous guidance or monitoring (N=8) are in the manufacturing domain (N=4), with examples as Hirt et al. (2019) where a physical instructor is involved in the tutorial phase and task phase and Schwarz et al. (2020) reporting a trainer's involvement to operate the control user interface, to instruct the participant, to start and monitor training and also to troubleshoot problems. Other cases, like Koumaditis et al. (2020b), in their assembly training task had partially utilised a trainer's verbal assistance as one of the dependent variables, and in Murcia-Lopez and Steed (2018)'s experiment, some guidance to the participants and physical management of cables is mentioned. In other domains, Liang et al. (2019) in their safety training scenario for the mining industry included a virtual instructor in the IVR environment, but a real instructor was also present to watch the live video feed of the IVR activities and to give instructions. In the emergency services sector, Nahavandi et al. (2019) in their firefighting training simulation have instructors monitor the breathing of users (recorded through pressure sensors).

Remote training is represented only in three industry domains. In more detail, manufacturing (N=2) has one publication with remote peer collaboration and one with remote guidance by an expert. In the first case, Wang et al. (2019) created a remote training setup with the trainee learning an IVR-based assembly task with the trainer giving instructions remotely via video/audio projected into the IVR environment, while in the case of Yildiz et al. (2019), two remotely located workers (both wearing VR HMDs) were collaboratively learning an IVR-based assembly operation. In the defence domain, Duggan et al. (2019) incorporate a remote trainer to observe and instruct trainees with the help of a mixed reality HMD, while the trainees themselves are present in an IVR environment wearing a VR HMD. Additionally, in the education services, Simeone et al. (2019) include an expert trainer to teach concepts of Virtual Reality to a trainee, where both users are inside the IVR environment represented by avatars (blue orbs).

Trends in hardware complexity

More than half the publications (N=46) use standard VR kit hardware, followed by a third of publications (N=27) using at least one additional device and only five using multiple devices alongside the VR kit, as seen in Figure 28. For clarification, it is worth noting, that the categorisation in this section is not a replication of previous sections (e.g. haptics or sensors) as there are cases here of portable haptics, for example, which are classified as part of standard VR kits (as they are essentially using the controllers associated with the HMDs) and there are cases of additional hardware that fall outside of the bounds of haptic devices and biosensors (for example handheld Xbox controllers and motion tracking equipment like the Kinect) (see definitions in tables 7, 8 and 11).

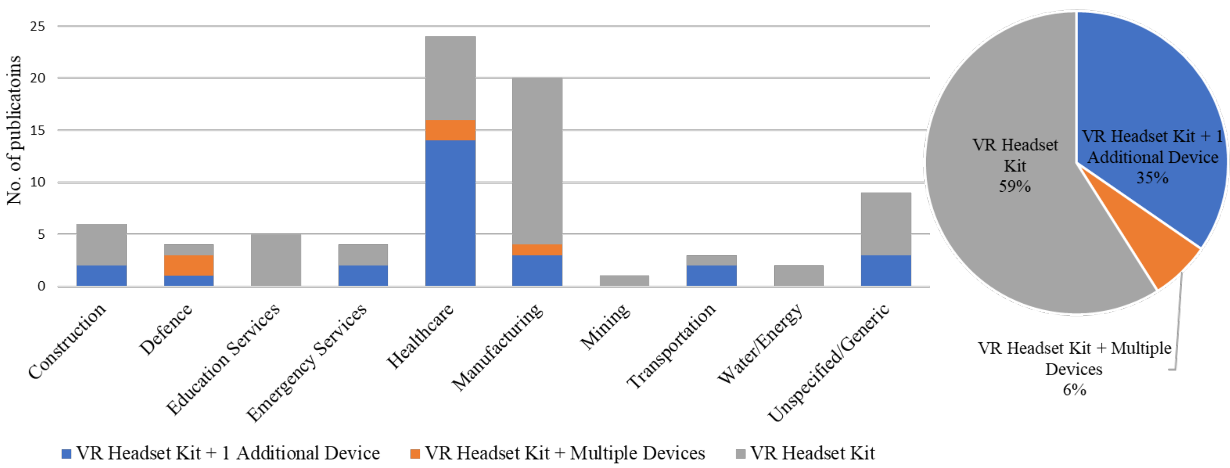


Figure 28 Trends in hardware complexity across publications and industries

Some illustrative examples of the latter include (Pulijala et al., 2018) using a Leap Motion sensor for finger tracking and (Akanmu et al., 2020) using a PrioVR motion tracking suit for training postural skills. Additionally, in some cases custom devices are utilised alongside standard VR hardware; for example, in the healthcare sector, Pelliccia et al. (2020) report the use of a modified KUKA industrial robot arm along with an HTC VIVE, and Durai et al. (2019) mention the use of a custom grounded haptic device for CPR training skills. Similarly, Nahavandi et al. (2019) in the emergency services domain depict a custom haptic device for firefighting.

Additionally, one may observe that publications in the healthcare (N=2), defence (N=2) and manufacturing (N=1) domains use multiple devices alongside the standard VR kit. In the defence sector, Muñoz et al. (2019) used two devices alongside the HTC VIVE Pro, a Polar H-10 chest strap for measuring heart rate and a Muse BCI headband for measuring EEG signals. While in the healthcare sector, Sainsbury et al. (2020) utilise a Leap Motion sensor along with a grounded haptics device for surgical training. In a bimanual assembly task training scenario in the manufacturing sector, Carlson et al. (2015) use a 5DT glove and an Immersion Phantom Touch device alongside an unspecified VR HMD.

Analysing industry trends, one may observe that emergency services (N=5), mining (N=1), and water/energy (N=2) depend only on VR headset kits, while the majority of papers in the healthcare domain (N=14) use at least one additional device alongside the VR headset kit.

2.1.6 Discussion

2.1.6.1 Analysis of results

Is IVR an effective training method for industrial skills training?

Indeed, according to our review, IVR can be an effective training method for industrial skills training. One-third of the publications in our database (N=29) reported IVR training to be effective compared to other VR modalities (CAVE, desktop VR, physical training, etc.), while only a small number of studies (N=6) found that IVR training was inferior but still capable to provide training in some extent. None of the studies indicated IVR to be unsuitable for training industrial workers. Among publications (N=6) that found IVR to be less effective than other modalities, we see that half the papers are from the manufacturing domain. Yet it is worth mentioning that a clear, detailed recipe of what enhances the effectiveness of IVR training cannot be reported. However, some indications of the impact made by design choices and user perception exist. User perception includes factors like perceived usefulness, self-efficacy, engagement and enjoyment which has been indicated to increase IVR effectiveness for example in: (Butt et al., 2018), (Lackey et al., 2016), (Allcoat & von Mühlenen, 2018), (Meyer et al., 2019), (Nykänen et al., 2020), (Pulijala et al., 2018), (Wu et al., 2020). Whereas, regarding design choices affecting IVR effectiveness, one may observe that the design of the IVR environment and user interactions (e.g.: head rotation amplification, restriction to arm-scale interaction) may negatively affect training performance to varying degrees. The latter was observed in: (Frederiksen et al., 2020), (Huber et al., 2017), (Ragan et al., 2017), (Sun et al., 2019), (Srivastava et al., 2019). On the other hand, certain design choices for increasing IVR realism may improve training success in IVR, for example with better object interaction technologies (haptic feedback, motion tracking, etc.) as was depicted in: (Winther et al., 2020), (Koumaditis et al., 2020b), (Zhou et al., 2019), (Mirauda et al., 2020), (Xin et al., 2020), virtual body ownership as seen in Škola et al. (2019) and other features which enhance immersion like visual fidelity found in Lang et al. (2018) and Zinchenko et al. (2020). Supplementary to the above one may revisit Section 2.1.5.5's subsection "Design attributes and effectiveness" for a more detailed analysis.

Moving to the categories of skills that the user of an IVR training system can acquire, it appears that procedural skills with 45% (N=35) and perceptual-motor skills with 33% (N=26) were the most taught types. Also, decision-making with 17% (N=13) and spatial skills with 5% of the papers (N=5) were reported in the body of literature, with a smaller number of representations, however. This portrays IVR systems as capable, to some extent, to be used as a training tool for any kind of industrial skills training, from processes and assemblies through object and tool manipulation to critical thinking. In more detail, regarding the industries and skill types, of the 35 publications focusing on procedural skills, half of the publications (N=12) were from the manufacturing sector. Quite a reasonable result, as this sector has an extensive list of training requirements that range from assembly tasks to maintenance and health and safety procedures. Expected results were observed in healthcare, a highly process-oriented field with specialised motor skill training needs. In our investigation, healthcare holds the majority of perceptual-motor skills (N=19), with a focus on the training of fine or gross motor skills in surgical tasks. The remaining industrial categories contain various skill types in relatively small percentages. An interesting observation comes from the small percentage of motor skills accounted for in the manufacturing sector (N=1). Based on our results, we assume that manufacturing has not yet applied motor skills training with IVR systems, or at least to the degree that it was applied in healthcare. Possibly because IVR motor skills training simulations require interweaved methods and tools, i.e., VR and haptics. This is evident in the results that portray healthcare, a highly mature domain in haptic training simulations, as the leader in the IVR motor skills training.

Additionally, the average number of participants per experimental condition per industry was presented. Upon inspection, the emergency services sector leads with an average of 75.75 participants per condition but it only has a small number of studies (N=4), followed by the education sector with 22.9 participants (N=5), and the water/energy sector with 29.5 participants, again with only two studies. Healthcare, which holds a significant number of studies (N=24), has on average 16.54 participants. While one cannot dispute the validity of the peer-reviewed studies presented herein, neither their findings based in most cases on statistically significant results, we as IVR researchers need to reflect if there is a need to augment the power of our contribution by increasing the size of participants. Brysbaert (2019) addresses the issue and points to studies with statistical significance but low numbers of participants, for which he finds drawbacks that can be lifted with the increase of the sample size per condition in the hundreds. This understandably might impose time, resources, and cost challenges; yet, strengthening the confidence of the reported results might also increase the applicability of the IVR training method and anchor it as a validated training method.

In addition to the number of participants, a trend worth mentioning is the time of the testing/assessment (immediately or after a period). The findings are quite revealing and expose a clear tendency. Of the publications measuring the effectiveness of IVR (N=40), we found that, although they all immediately measured the results of the training (skill acquisition, knowledge gain, enjoyment, etc.), only nine papers measured the effectiveness of IVR over longer periods, with relatively positive results. Thus, a clear trend of immediate testing exists. This is not always the case if one investigates beyond our data set. Examples of longitudinal studies appear every so often in literature (e.g., VR for diabetes (Vorderstrasse et al., 2015) or Parkinson's and VR (Mendes et al., 2012)), but to the best of our knowledge, these kinds of longitudinal studies are predominantly healthcare clinical studies and not IVR in industrial skills training. The contribution of a longitudinal study in the IVR industrial training might address the effect of memory/skill retention after a considerate period in time. For example, the Ebbinghaus model (Ebbinghaus, 2013) suggests that students forget around 75% of what they learned within a few days in non-VR studies. Similarly, White and Arzi (2005) suggest that learning, in general, can be an erratic process, so there is a chance that a learning effect may not be immediately apparent resulting in a negative result, which may change if the test is conducted at a later date. On the other hand, there is also a chance that the learning effect is short-term and volatile, and a positive effect may not remain when tested later. Thus, one might consider that further research is required, inclusive of the long-term effects of IVR training, over time.

How do we conduct research in the IVR for the industrial training field?

It is observed that 62% of the publications (N=48) have an arm-scale VR training space, and the user is either seated or standing (without moving around). Another 36% use a room-scale space (N=28) where different tracking technologies enable tracked navigation around a small area. The reason that one might favour an arm-scale interaction space is mainly attributed to the requirements of the task itself. Our hypothesis is that the technology (tethered HMDs, grounded haptics, etc.) and/or the need to minimise the cybersickness effect might contribute as well. Additionally, virtual environments provide opportunities to extend the virtual area of interaction without the need for an extensive physical space. One example is teleportation, seen in our database (Allcoat & von Mühlenen, 2018; Duggan et al., 2019; He et al., 2019; Hirt et al., 2019; Muñoz et al., 2019; Pérez et al., 2019; Shamsuzzoha et al., 2019; Simeone et al., 2019; Zhao et al., 2019). Examples of newer techniques like those by Feuchtner and Müller (2017) using elongated hand interactions may be employed to extend the effective area of interaction and thus accommodate for a disparity between physical and virtual space. Possibilities also exist for extending the range of walking in both arm-scale and constrained room-scale in IVR environments by redirected walking and the provision of different kinds of stimuli (visual, audio and haptic feedback) as detailed by Nilsson et al. in their review of this field Nilsson et al. (2018).

As for the display technologies used in IVR training in the industry, the most widely used HMD was the HTC VIVE with 45% of the papers (N=35) reporting on its use, followed by the Oculus Rift employed by 26% of the papers (N=20). Eight papers use the slightly more advanced version of the HTC VIVE – the VIVE Pro. As expected, the Oculus Quest HMD as a new arrival was not observed in the data set; yet, its untethered features and low cost might make it a highly popular alternative in the years to come and may affect other trends observed herein (i.e. the VR training space).

One may think that IVR systems, to a high degree, are designed based on applied, known learning theories. This, however, is not evident in our data. Only in 14% (N=11) of the total publications did the authors explicitly mention a learning theory as a design parameter. Similar results are reported by other VR reviews; for example, Radianti et al. found only a third of the papers in their review explicitly mention taking inspiration from a learning theory (Radianti et al., 2020). In our case, the papers that reported a learning theory (N=14), 13 examples were identified, constituting a comprehensive, valuable list for IVR designers and developers (depicted in section 2.1.5.2). Though most of the learning theories we found in the 11 papers are not specific to the medium of IVR and can be generalised to any medium, the use of pre-training principle and the cognitive theory of multimedia learning as employed by Meyer et al. shows the potential of the interaction of media (IVR) and the training method.

Still, on the IVR design considerations, the indications are stronger when focusing on the experimental design method. 82% of the publications used an experimental design approach (N=64), including user surveys and between-/within-subjects studies, followed by 17% of the publications focusing on the design and development of IVR training without an emphasis on validation (N=13). In more detail, the between-subjects validation method had the highest representation at 62% (N=48), followed by the within-subjects and preliminary study/survey with a share of 15% (N=12) each. Six publications did not mention any kind of validation (no study). Unsurprisingly, in papers using between-subjects validation, the healthcare domain had the highest representation with 16 papers, followed by manufacturing with 13. Understandably, the aforementioned findings do not expose any new unique approaches, however, they depict the academic foundation that IVR research stands upon. A foundation further analysed in the following sections, in terms of measures like time, scores and immersion.

In our data analysis, the most popular measure was time, with 42% (N=33) of the papers measuring task completion time and five measuring reaction time. The former were mainly clustered in the manufacturing (N=14) and healthcare (N=9) sectors. The second-most popular parameter after time was variations of scoring metrics. These variations included keeping track of the number of errors (N=11), successful steps completed or accuracy (N=20) and other measures like performance in knowledge retention tests (N=10). In more detail, the healthcare domain used scoring metrics the most with ten publications, followed by manufacturing with six, education services with two and one from each of the other domains with four publications from an unspecified/generic domain. As a clear consensus on the effectiveness of IVR in the field does not exist, one may expect these high percentages of basic measurements of time and errors or accuracy to dominate the findings.

In contrast to the high percentages of time and scores, small percentages of measurement of parameters like cybersickness with 18% (N=14), task load with 17% (N=13) and immersion with 18% (N=14) were identified. While the low percentages in cybersickness measurements might be expected due to a) the number of high-fidelity HMDs utilised and b) the number of papers that minimised that effect by placing the trainees in a sitting position (N=28), the same cannot be said for the task load and immersion measurements. Task load measurements may be correlated with theoretical methods of cognitive load such as *cognitive load theory* and *cognitive theory of multimedia learning,* both identified in our data set with small numbers of publications (N=2 and N=1, respectively). But then again, the 'one', i.e. lack of task load measures, is a consequence of the 'other', i.e. lack of task load theories, and not a justification for the small percentages. What is more, literature beyond IVR highlights the importance of task load in the design and evaluation of instructional technology-based training (Brünken et al., 2003)? Why, then, is this absent from our data? Possible explanations could be that the field is not yet mature enough to apply such conditions due to the complexity of the task, the lack of tools and methods and the validity of the subjective data (the most common type of data gathered).

Immersion was another parameter that was underrepresented in the data set. This is worthy of reflection, especially if taking into account literature beyond our findings. For example, the effect of immersion on memory and presence in virtual environments has been highlighted in Mania and Chalmers (2001) who found it to be a compelling research parameter. In fact, in other fields, e.g., rehabilitation, the lack of conclusive evidence for the link between immersion and outcome has been pointed out (Rose et al., 2018).

Additionally, a variety of data analysis methods, including descriptive statistics (N=19), ANOVA tests (N=24) and t-tests (N=13), were significantly represented. All the identified data analysis methods are well established and justifiable elements of an IVR study.

How can we improve IVR training for the industry and make it applicable for remote training?

Understandably, IVR training is a unique training method. The trainee is immersed in a virtual, controlled, interactive environment with the opportunity to have a quantifiable real-time snapshot of his/her bio-signals during the VR training experience. Inspirations from non-VR training literature include Sutarto et al. (2010) where the authors describe the use of heart rate variability (HRV) based biofeedback training for industrial operators. However, this opportunity to use biosensors was not extensively evident in our data set, as only 14% (N=11) of the publications used these data-gathering devices. Still, those that did measure brain signals (N=6), heart rate (N=2), eye tracking (N=4), temperature (N=1) and skin conductance (N=1). None of these papers used any form of haptic interaction. Even though the small number of cases prevents any generalisation, it is worth mentioning that a balanced spread of biosensors across industrial domains, skills and location on the body was observed.

32% of the publications (N=25) used a variant of haptic feedback. Of these papers utilising haptics, almost half (N=11) used grounded haptics, six used portable haptics, five used pseudo-haptics and three used wearable haptics. Unsurprisingly, almost two-thirds of these papers (N=15) belonged to the healthcare domain, a domain pioneering haptic use cases long before they were applied for training with VR. For example, in the surgical training domain, haptic simulation is one of the most mature use cases one can find in the VR literature, with decades of studies. In the rest of the industrial categories, the manufacturing and defence sectors hold two papers, respectively, while the remaining industrial categories hold a paper each. A trend identified in the data is that, from the 25 papers using haptics, 80% (N=20) are arm-scale cases and half of these apply grounded haptics. This makes sense, as haptic devices and especially grounded haptics are complex systems that require maintenance and systematic calibration; thus, stabilising the device in a fixed grounded position in arm-scale is reasonable. Arm-scaled grounded devices require easier maintenance operations as compared to room-scaled devices. This has probably played in favour of their presence in literature cases. Moreover, as the technology is not limited by portability/wearability capabilities, they can rely upon designs aimed at providing reliable and accurate performance, making them quite adaptable for the above-mentioned healthcare cases. In the case of the applicability of haptics in remote training use cases, it is evident that challenges and limitations exist; for example, the utilisation of grounded haptic devices in surgical training like the Kuka industrial robot arm in Pelliccia et al. (2020) or the Lapmentor III platform (Li et al., 2020) might be challenging to be incorporated into a remote scheme for training, but as the technology matures, other simulation techniques like haptic gloves that were observed in Butt et al. (2018) can aid to partially or fully transfer the IVR applications into a remote training scheme.

The need for IVR to be applied as remote training has increased due to the COVID-19 pandemic and the physical distancing restrictions applied to work and training settings. To do so, IVR applications need to be examined based on the trainee-trainer interaction and technical requirements. On the issue, the review findings indicate most of the IVR applications to be applicable for remote training, either classified as autonomous IVR training (N=60) or designed with remote human interactions (N=4). In more detail, three publications mention a remote trainer (remote guidance) and one requiring peer learners located remotely (remote peer collaboration). This is a positive and expected finding, as IVR training, due to the digitalisation of the training process, is usually designed to be applied remotely. Thus, a lot of solutions can be applied for remote training under the condition that the trainee has access to the required VR kit and the digital content. These practices were identified in examples like Li et al. (2020) where a virtual instructor instructs the trainee inside the IVR environment throughout the training and Liang et al. (Liang et al., 2019) where a virtual instructor provides the introduction, training overview and instructions to the trainee in rock-related hazard safety training. Our findings highlight that in most cases, a standard VR kit hardware (N=46) was utilised; thus, the aforementioned models can be applied. In the COVID-19 era, anecdotal cases utilising innovative business models where the training supplier also supplies the VR kit (typically headset and controllers) and lends it to the trainees have emerged (Grensing-Pophal, 2020; Ivec, 2020). However, if these emerging possibilities will reposition IVR training in a higher position after the COVID-19 era, it remains to be seen.

2.1.6.2 Limitations

This review holds several limitations that must be considered when applying the study's findings. First, the studies analysed were identified based on our selection criteria and focused on peer-review publications. Thus, as in any systematic review, some significant knowledge and results might have been filtered out. For example, information from industry reports, books and cases could provide additional insights. Along the same lines, our choice of categorisation (inspired by the well-known Global Industry Classification Standard (GICS), but supplemented with additional categories), can be considered as a necessary but restrictive approach. Second, our definition of an IVR user/trainee includes a healthy neurotypical adult with no profound mental/physical disabilities or illness. Our view is that IVR training systems need to be inclusive of workers with all abilities, mental and physical, always, of course, aligned with health and safety restrictions. However, we decided to define the user as we did to grasp a more generic, representative body of literature and limit our literature count. Future research could go beyond these limitations and extend our findings by conducting replicative studies in different contexts with different technological applications.

2.1.7 Future Research Directions

This systematic review highlighted several converging themes which inform future research directions in IVR training, and which have been grouped below for the benefit of researchers seeking inspiration for extending the field:

- *Theoretical grounding:* As discussed in the previous sections, a relatively small percentage of only 14% of the representative papers exposed a learning theory as part of their conceptualisation and design. One might hypothesise that a stronger grounding might create more effective training, or at least expose possible incompatibilities of IVR's nature and the utilised learning theory or even the need for a more tailored approach. To this end, the Cognitive Affective Model of Immersive Learning (CAMIL) by Makransky and Petersen (2021), which discusses the need to leverage the unique affordances given to learners by the IVR medium, namely presence and agency. This approach might be indicative of the way that learning theories and IVR can be addressed and such cohesion can form a future direction for research.
- *Longitudinal studies*: A clear trend of immediate testing was depicted in the data gathered in this paper, whereas, only nine papers measured the effectiveness of IVR over longer periods. Understandably, IVR effectiveness for industrial training applications is not yet widely researched. Interesting cases of longitudinal studies exist like VR for diabetes or Parkinson's and VR, but to the best of our knowledge, these kinds of longitudinal studies are predominantly healthcare clinical studies and not IVR in industrial skills training.
- *Adaptation based on bio-signals:* One might argue that IVR training's novelty resonates in the digitalisation of the training experience and the ability to quantify the trainees' interactions and physiological signals, during such experience. Herein, Muñoz et al. (2019), Dey et al. (2019) and Škola et al. (2019) grasp this offering, utilised real-time bio-signals and created a layer of adaptiveness in the IVR training. It is our understanding that such adaptive IVR applications, their design, architecture, and attributes, can open up an exciting new research path for the future.

- *Remote training:* The potential for remote IVR training is promising and timely considering the challenges and changes in industrial training brought in by the COVID-19 pandemic. The review findings depicted most of the IVR applications to be applicable for remote training. This is a positive but expected finding, as IVR training, due to the digitalisation of the training process, is usually designed to be applied remotely. However, the business models that can sustain IVR training solutions are unexplored and can be of potential interest to the IVR community. Such research might expose the interplay of architecture, design, and socio-economic issues.

2.1.8 Conclusion

In this paper, a systematic review was conducted that focused on the application of immersive VR technologies for industrial skills training. The investigation of immersive VR technologies included, among others: application categories, design parameters, learning theories, data analysis methods, infrastructure, sensors, haptics, and potential for remote applicability.

The review revealed a growing interest in IVR training from various industrial domains, ranging from manufacturing and assembly training to service providers in healthcare and defence. The results portray IVR as an effective training method capable of transferring procedural, decision-making, spatial and fine/gross motor skills. Yet, how this is achieved and what the parameters are that influence the effectiveness remains unexplored. Most of the cases treat IVR as a new paradigm that is in an experimental state and mostly assess it in terms of effectiveness, i.e., time, usability and scores.

Among the domains investigated, healthcare surfaced as a champion in terms of the number of studies, use of haptics and effectiveness of the IVR training. Thus, when studying the publications, we identified in the healthcare domain that a person could be exposed to successful, efficient IVR systems that ground IVR as a valid training method. However, in healthcare, very few studies designed their VR applications based on a specific learning theory. To a great extent, the same is true for most industrial domains. Adding this limitation to the small number of identified studies with effectiveness measures (additional to time, usability, and scores) makes it challenging to extract the parameters or combination of parameters that made the IVR effective. Parameters like immersion, memory and presence require further investigation and specialised, focused studies. Still, we did highlight several design attributes that seem to play a crucial role in the effectiveness: object interaction realism, immersive features, break of presence, virtual body ownership and gamification. One needs to reflect on these attributes, not as basic ingredients in a recipe for success, but as focal points of an uncharted framework.

Our review also exposed unexplored but intriguing avenues of research, like the utilisation of biosensors for users' data collection and the applicability of IVR applications as remote training solutions. To the latter, we highlighted categories that can potentially, with the integration of devices, redesign of trainers' interaction or new business models, create IVR solutions for remote training.

2.2 Immersive Virtual Reality Training: Three Cases from the Danish Industry

Effective industrial training has always been vital. Recently, the need for robust, safe, repeatable, and cost-effective digital applications drove many industries to explore immersive technologies and especially Immersive Virtual Reality (IVR) as a possible solution. In this paper, we depict three such cases, from the Danish industry and showcase the motivation of such adoption, technological and design characteristics, alongside perception of its applicability.

2.2.1 Introduction

On average, organizations spent 10% of their budget on learning tools and technologies with the most frequent purchases on e-learning and systems, learning management systems, and simulations (Freifeld, 2018). These investments in training activities allow organizations to adapt, compete, excel, innovate, produce, be safe, improve services, and reach goals (Grossman & Salas, 2011). Training is effectively deployed to decrease errors in such high-risk work settings as emergency rooms, aviation, and the military.

Now, a new innovative wave of Immersive Technologies (Virtual/Augmented/Mixed Reality - VR/AR/MR) is being utilized as training tools. The need for such implementations reflects the requirements for cost-effective, safe, scalable, modular and mobile systems. Yet, this is a new phenomenon, with unexplored issues such as the motivation for adoption, trends in design and technologies, assessment techniques, and perception of stakeholders.

- RQ1: What motivates industries to adopt IVR training?
- RQ2: What technological parameters and design elements are incorporated in the IVR training?

- RQ3: How is IVR training assessed and perceived by the stakeholders?

The rest of the paper is structured as follows: Section 2.2.2 provides a short introduction to the theoretical background, Section 2.2.3 depicts the methodology, Section 2.2.4 the case organizations and the IVR applications, Section 2.2.5 the analysis and discussion, and Section 2.2.6 the conclusions.

2.2.2 Theoretical background

Effective industrial training has always been paramount. Literature, reports that the cost for fail-to-recall procedural tasks in a production environment is high, and errors can be catastrophic for the product and the overall production cost (A. C. Falck et al., 2010). Similar shortcomings apply when errors occur beyond the production line, for example, assembly of machinery or installation of electronic components in the field of business or hard to approach locations, like offshore wind turbines or oil drilling rigs (Reason & Hobbs, 2017). However, training is not as intuitive as it may seem. Literature portrays variations in the way one can design, deliver, and implement a training program (Salas et al., 2012).

Thus, it comes as no surprise to observe a great interest and investment from industry and academia on new innovative approaches to enhance training procedures. To this a new set of immersive technologies as Virtual Reality (VR); a fully digital environment, Augmented Reality (AR); an environment where the digital content augments the real world and Mixed Reality (MR); the combination of real and digital with interactive elements, has emerged to provide compelling solutions. Immersive technologies have been successfully utilized for example in health and safety (Ayala García et al., 2016; Grabowski & Jankowski, 2015), disaster preparedness and response training (Hsu et al., 2013; Pucher et al., 2014), surgery training (Gurusamy et al., 2009; Seymour et al., 2002) and assembly tasks (Carlson et al., 2015; Koumaditis et al., 2020b).

Currently, Immersive Technologies are in a state of continuous development, with new affordable head-mounted displays for VR and AR glasses, advanced untethered communication capabilities, and increasing maturity in software development tools (Anthes et al., 2016) and assessment techniques (Moore et al., 2020). These technologies if utilized correctly can facilitate a risk-reduced, innovative, robust, cost-effective, repeatable practice environment (Koumaditis et al., 2018; Okuma & Kurata, 2016).

The Danish industry sector has welcomed this opportunity, and in recent years a business ecosystem of VR developing companies (mainly SMEs) and big industrial partners (mainly in the renewable and services sectors) has been created. In this setting, the investigation described herein took place.

2.2.3 Methodology

The focus in Denmark was placed mainly due to the suitability to explore this phenomenon (availability of IVR cases) and the constrained travel due to the COVID19 restrictions that placed a barrier to explore cases in other countries. The three cases mentioned herein are the shortlist of many inspiring examples of IVR training we came across but these were chosen both for their representability of the IVR cases but also for the availability of case data.

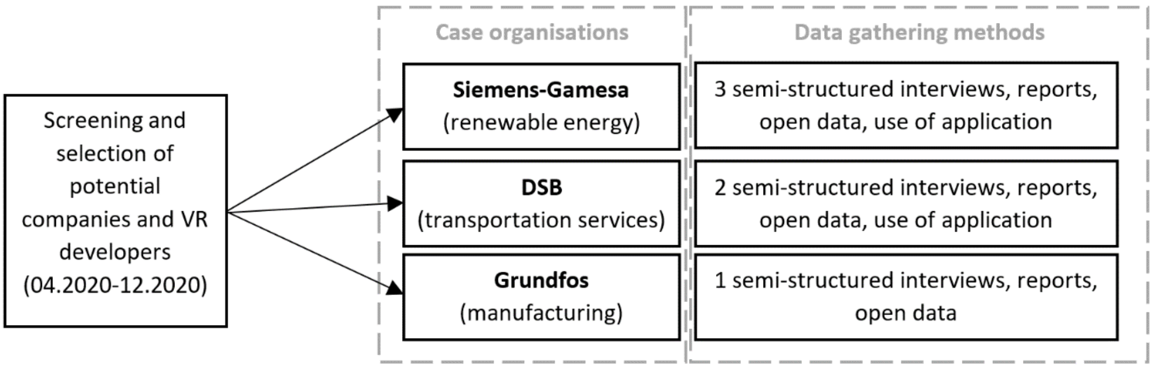


Figure 29 Case organizations and data gathering methods.

In order to answer the three aforementioned research questions a thorough inspection of each case was performed including the gathering and analysis of a) semi-structured interviews, b) inspection of the IVR application and field studies, and c) open published data and available reports, as seen in Fig. 29.

2.2.4 Case organizations

2.2.4.1 Siemens Gamesa

Siemens Gamesa[11] is a Spanish-German wind engineering company, which manufactures wind turbines and provides onshore and offshore wind services. Siemens-Gamesa is a key player and innovative pioneer in the renewable energy sector; they have installed products and technology in more than 90 countries. It is the world's second-largest wind turbine manufacturer.

[11] https://www.siemensgamesa.com/en-int

Challenge: Part of their offshore wind turbine installations is to train workers into performing specialised installations (referred to as "turning tool" training). Reported challenges of such training are the weather conditions and health and safety regulation when the training takes place in the offshore turbine in the sea and cost (renting the turning tool and transporting trainees and trainers) when the training is conducted in a simulated environment in the land.

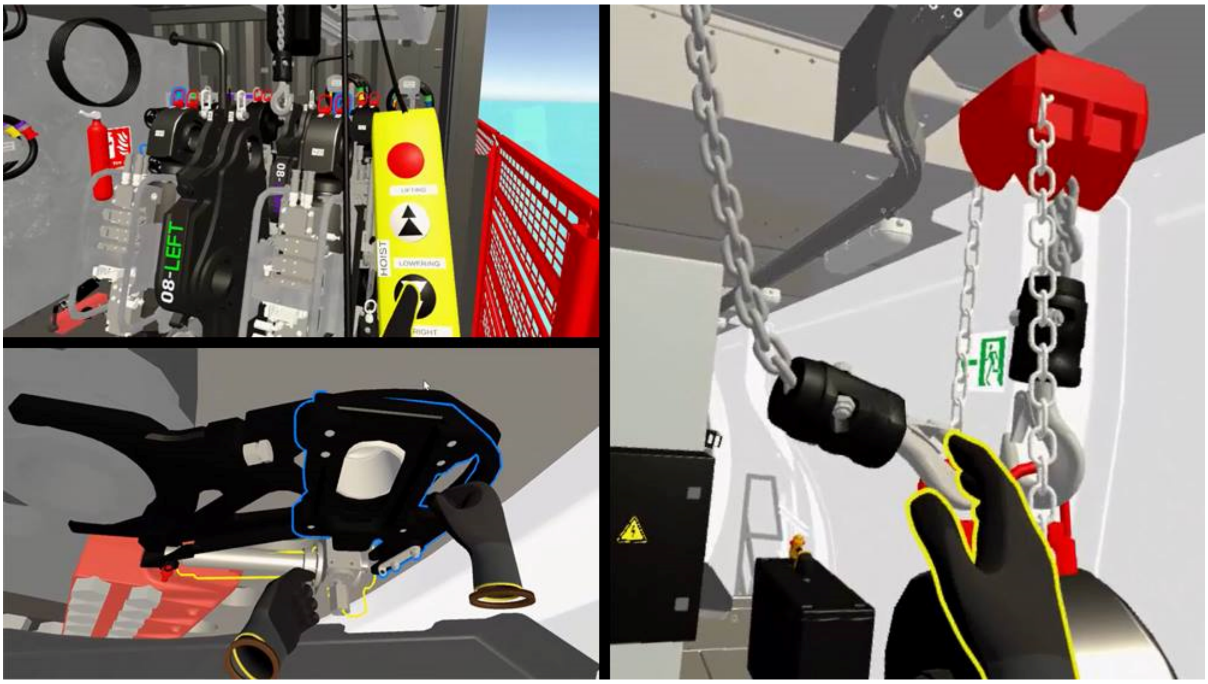


Figure 30 IVR training solution in Siemens-Gamesa.

Motivation: One of the main objectives for the Siemens-Gamesa training department to adopt IVR training was to provide an easy to transport, safe, and repeatable training solution that will reduce the cost of training for technicians involved in the installation of wind turbines. The development of the IVR training was carried out by Kanda[12], a Danish VR development company and was implemented in an iterative proof-of-concept process where funds were allocated as the solution matured.

Skills: The skills that are taught with the IVR solution primarily involve procedural skills and familiarisation with the work environment and tools. This includes process training steps on how to assemble parts of the wind turbine, how to identify areas where objects like screws are placed and more complex skills like operating a telescopic crane among others. In

[12] https://www.kanda.dk/

the case of the telescopic crane, technicians need to learn the use of an internal container telescopic crane to lower down equipment from the container on the top of the wind turbine nacelle.

Design Elements: Gamification and multi-user interactivity are integral parts of this IVR solution and has been designed to allow people to make mistakes. Time is not measured in any manner. Additionally, visual aids are present in the IVR training environment to guide the learner. For example, tools - objects that needed to be utilised to complete tasks were highlighted, as well as the areas where they were to be placed. The environment is not photorealistic but feedback from trainees and trainers suggest that the virtual environment is highly realistic, as seen in Fig. 30.

Technological Elements:

- Virtual Reality – The HTC Vive[13] is used for the currently deployed VR training. For most parts of the training, the interactions involve two hands (using the VR controllers). The workspace is room-scale, as the trainees have to kneel down, walk around and turn. For further movement, they teleport around the VR environment.
- Haptics - Haptic feedback is used in the simulator to give vibrations when trainees walk into virtual walls. Still, there are no vibrations while interacting with objects. The interviewees however indicated that conveying weight and texture information during interaction with virtual objects would be helpful for skill training. Physical controllers were not used in order to reduce costs.
- User Adaptation – An interesting observation from the training sessions revealed that trainees with experience playing video games adapted to the training faster and seemed more enthusiastic in learning the VR interaction features and the use of the controller. On the other hand, there was not any significant difference between the aforementioned users and the rest of the trained workers in terms of learning performance. Nevertheless, the company expressed interest in adapting the IVR training aligning dissimilar trainee profiles.

Assessments: There are no assessments, or any kind of score-tracking features incorporated with the IVR experience. Instead, they rely on the trainer's observations to analyse

[13] https://www.vive.com/us/product/#pro%20series

and decide on whether the trainees are learning or not, mainly based on the progress observed. The trainers view the trainee's activity and progress through a desktop monitor.

Outcome: The interviewees reported on the positive outcome of the application of IVR as a training tool and highlighted that has been rolled out in several training settings. However, as no metrics, comparison or statistics are available it was not clear if the IVR training has been more effective than traditional methods, but a documented benefit was that the time for renting out the turning tool equipment had been reduced from two weeks to one. Additionally, and mainly through the perceived notion of it beneficial use the investment and the IVR as a training technology was seen with a positive eye and considered as worth exploring further.

2.2.4.2 DSB

DSB[14] is the largest Danish train operating company and the largest in Scandinavia. As of 2005, DSB employs about 9,000 people with an increasing number of services each year.

Figure 31 IVR training solution in DSB.

Challenge: The company employs various roles to service, in the train, during a journey, amongst them the train operators. Part of the train operators' job is to manage, in a timely manner, opening and closing of doors, when the trains are stationary at stations and or in the case of an emergency. Training takes place by decommissioning the train and providing access to the trainer and trainees to proceed with the training scenarios.

[14] https://www.dsb.dk/om-dsb/

Motivation: The aforementioned method of training causes unwanted downtime in functioning trains and does not always provide the realistic scenario i.e. that of a functioning, noisy station is resembling real life. Thus, an IVR experience that may simulate such a scenario was created. The key objective is to train DSB train operators to learn the use of the whistle and door operation.

Design Elements: The trainees learn the procedures related to managing the opening and closing of train doors in a linear single-user IVR experience, (no decision making, or fine motor skills are taught). The virtual environment is a detailed representation of the train station platform, the train doors, and the immediate interior of the train coach behind the doors. The trainees navigate the virtual environment through teleportation while standing in one physical location. Visual cues, as seen in Fig. 31 are present to aid the trainee by displaying time and highlight points-of-interest. The trainees first undertake a theory class on the subject matter before the IVR experience. Training is delivered at the headquarters in batches of 12 trainees, who are split into groups of two. Each group is assigned an IVR headset and a tablet. There are two expert trainers who manage the training experience and aid the trainees reflect on what they learned.

Technological Elements:

- Virtual Reality – The trainee uses an Oculus Quest and an iPad, where one trainee experiences IVR training in the Quest while the other observes the progression and provides feedback.
- Haptics - Haptic feedback is present as an indicator of the trainee that certain actions were performed. For example, when a trainee opens the train doors gets vibration feedback to indicate successful completion. This is meant more like a nudge than to represent a physically realistic interaction.
- User Adaptation – No adaptation for specific users was identified.

Assessments: There are no formal assessments for the IVR training. Completion time is logged and provided in the trainees, as well as the oral peer-feedback they get during the training, but neither is analysed or compared with other relative data.

Outcome: The interviewees perceived the IVR as a beneficial training method, still requiring development but due to the ever-increasing demand for fast training at a lower cost a step towards the right direction.

2.2.4.3 Grundfos

Grundfos[15] is the largest pump manufacturer in the world, with more than 18,000 employees globally. The annual production of more than 16 million pump units, circulator pumps, submersible pumps, and centrifugal pumps.

Figure 32 IVR training solution in Grundfos.

Challenge: Training employees in machine operation and safety procedures during the manufacturing process is a crucial task that benefits error and accident reduction. For such training machines are diverted from production to an idle state, increasing downtime and the need for maintenance and recalibration.

Motivation: The VR training was also intended to reduce downtime and maintenance of machines while providing a safe, global, reusable tool for training. The development of the IVR training was carried out by Unity Studios[16], a Danish VR development company.

Design Elements: The VR platform includes seven different training applications, most of it targeting production workers, where training consist of completing a series of virtual versions of physical installations of machines, such as montage, measures of depth and centering, assembling construction parts, and packing a completed pump and preparing it for transportation. The IVR scenarios included a realistic representation of the machines and environment, as seen in Fig. 32, and visual cues to guide the user to perform the training tasks.

Technological Elements:

- Virtual Reality – The VR hardware in use was the HTC Vive tracked on a room-scale. The VR environment was designed to exactly replicate the look and feel of

[15] https://www.grundfos.com/
[16] Unity Studios is now SynergyXR - https://synergyxr.com/

real-world Grundfos factories. Interactions were through the Vive controllers using both their tracked positions and the controller buttons. The trainees get audio, visual, and text cues guiding them to complete the training steps.

- Haptics - No haptic feedback was reported.
- User Adaptation – No adaptation for specific users was identified.

Assessments: There was no systematic validation of the IVR training. An informal contest-type assessment between IVR trained and non-IVR trained employees provided equal performance metrics, depicting the validity of IVR as a training tool.

Outcome: The IVR training is in use and stakeholders testify that with the use of IVR training, they observed an increase in the motivation and decrease of the time required to train employees. In addition, the employment of IVR provided the possibility of a) increased collaboration, as several VR users could train together on the same content regardless of time and space and b) deployment of a uniform training method and scenarios on a global scale. Nevertheless, some side effects like cybersickness, in a group of workers with no previous game experience, was also reported.

2.2.5 Analysis and Discussion

In this section, an attempt to integrate the findings from the three cases and answer the research questions posed as part of this research is depicted.

What motivates industries to adopt IVR training?

It was evident from the cases that even big companies, leaders in their fields, as the ones investigated herein, decided to invest in IVR training, precisely when a strong business case was made. In more detail, considerable training cost, scarce resources, downtime was parameters that made the investment in IVR training reasonable. Yet, it was also evident that the investments were dictated by an iterative process with funds released as the IVR content and features matured and tested, usually resulting in a proof-of-concept application that was then disseminated to the rest of the organisation. Other parameters reported as motivational to invest and adopt IVR training include safety, reusability, standardisation, and mobility.

What technological parameters and design elements are incorporated in the IVR training?

Design and technological elements – We identified certain design elements during the analysis of the use cases, and we categorised them as follows –

Hardware - All of the use cases involved well-known commercial HMDs like the HTC Vive. Nevertheless, the interviewees expressed interest to invest in untethered solutions like the Oculus Quest due to its portability, low cost, and ease of use. No specialised haptic devices or other apparatus were utilised to augment the training experience.

Multi-user interactions - Clear intentions to apply peer cooperation in some form or another was reported. Siemens-Gamesa had a multi-user experience where two users had to cooperate to operate a crane and move equipment to the wind turbine nacelle. On the other hand, in the DSB case, a pairwise training strategy with peer-feedback was applied. This may reflect a trend in industrial training, as industrial activities are essentially cooperative in nature.

Immersive features - Immersive VR has certain features that make it distinct, including the angle of view, three-dimensional interaction, head, and body tracking and immersive stimuli. In two of the cases a high-fidelity HMD (HTC Vive) is used along with precise tracking of the head and the controllers which leads to immersive experiences, while in one the Oculus Quest is used which offers the same capabilities but with a slightly lesser screen resolution. Unfortunately, the direct effects of the immersion were not mapped against any other parameter and/or the resulting performance.

Object interaction realism - We broadly define object interaction realism as the fidelity of the interaction with digital objects in the virtual environment. In all three cases, there is a focus on high visual fidelity with near replicas of the environment, yet it was observed that the interaction with the digital objects is not realistic, identical to the real physical objects and tools. This is attributed to the design of the experience and the utilization of the generic VR controllers, which do not represent the dimensions, weight, and other physical characteristics of the objects and tools the trainees are being trained to use. Interestingly, in the DSB and Siemens-Gamesa cases, they used haptic feedback, with the latter including haptic feedback for signifying when the trainee hit walls and in the DSB case, to signify successful completion of tasks. The use cases did not reveal any use of haptics for motor skill training.

Break in immersion - One may refer to "break in immersion" as any factor that can divert the user to lose their perceived feeling of immersion in the virtual environment. The Siemens-Gamesa and Grundfos cases seem to involve immersion in the VR environment, with the limitations of lack of object realism. However, the DSB case contains communication of the immersed in VR trainee with a fellow trainee situated outside the IVR environment (who views the IVR scene via an iPad). One may argue that this interaction might break immersion, affecting the training performance.

How is IVR training assessed and perceived by the stakeholders?

Assessments in all the cases were oriented around either peer giving feedback to each other, or experts present at the site giving feedback to the trainees or unstructured contests. There were no structured automated assessments with objective and subjective measures. Nevertheless, the perception of these proof-of-concept applications was reported as positive across the board. Even so that in all the cases intentions to increase the investment and augment the percentage of IVR training applications were reported.

2.2.6 Conclusion

In this research, three prominent examples of IVR training in renewable energy, transportation services, and manufacturing sectors were depicted. Such industrial examples clearly provide helpful insights into the motivation, technological and design characteristics, and perception of IVR training from a business point of view.

Our analysis depicted the value of building a business case in order to invest in IVR training, the use of commercial HMDs, the need for multi-user interactions, immersive features, object interaction realism, and immersion. Also, the lack of dedicated haptics or other apparatus and assessment methods. Our future plans involve experimentation upon these findings in order to frame IVR training and provide clear guidelines that can augment the training performance.

Chapter 3 - Investigating the Effectiveness of Immersive VR Skill Training and its Link to Physiological Arousal

This chapter details the motivations, design, and analysis of a study using a fine motor skill training task in both VR and physical conditions. The objective of this between-subjects study was to a) investigate the effectiveness of immersive virtual reality for training participants in the "buzz-wire" fine motor skill task compared to physical training and b) investigate the link between participants' arousal with their improvements in task performance. Physiological arousal levels arising from electro-dermal activity (EDA) and Heart Rate Variability (HRV) data calculated from ECG (Electrocardiogram) were collected from 87 participants, randomly distributed across the two conditions. Results indicated that VR training is as good as, or even slightly better than, training in physical training in improving task performance. Moreover, the participants in the VR condition reported an increase in self-efficacy and immersion while marginally significant differences were observed in the presence and the temporal demand (retrieved from NASA-TLX measurements). Participants in the VR condition showed on average less arousal than those in the physical condition. Though correlation analyses between performance metrics and arousal levels did not depict any statistically significant results, a closer examination of EDA values revealed that participants with lower arousal levels during training, across conditions, demonstrated better improvements in performance than those with higher arousal. These findings demonstrate the effectiveness of VR in training and the potential of using arousal and training performance data for designing adaptive VR training systems. This paper also discusses implications for researchers who consider using biosensors and VR for motor skill experiments.

3.1. Introduction

Virtual reality (VR) based training is increasing in popularity and is being explored in recent years across domains like education (Radianti et al., 2020), rehabilitation (Howard, 2017), and various industries targeting adult learners (Abich et al., 2021; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021; Renganayagalu et al., 2021; Xie et al., 2021). VR-based skill training brings in several advantages like allowing learners to practice procedures safely and repeatedly with consistent feedback (Hamilton et al., 2021). For example, in a Cochrane meta-analysis of studies investigating the effectiveness of VR training in endoscopy skills, it was found that VR training was more effective than no training and as effective as physical training (Khan et al., 2019). The advantages of VR training are being further enhanced by the increasingly widespread availability of Immersive VR (IVR) technologies which make use of CAVE (Cave Automatic Virtual Environment) technologies or head-mounted displays (HMDs), offering high fidelity audio-visuals to the user (Makransky et al., 2019). The immersion and presence offered by IVR further enhance its effectiveness, particularly when the affordances of IVR are matched with the teaching/training method (Makransky & Petersen, 2021). It must be noted that IVR still has limitations in comparison to physical reality, to name a few in particular: differences in visual acuity, field of view, and the presence of cybersickness, the latter possibly linked to differences in vestibular response (Ashiri et al., 2020). As the evidence for the effectiveness of IVR over other methods is mixed (Abich et al., 2021; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021), one may ask: how can IVR training be improved?

IVR training primarily makes use of easily observable training/test performance metrics like task completion time and the number of errors (Abich et al., 2021; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021). In addition to such objective measures, literature on skill training outside of IVR has also investigated the links between arousal and performance (Storbeck & Clore, 2008; Yerkes & Dodson, 1908). The term *arousal* refers to many related phenomena like an increase in alertness, attention, emotion, or the ability to respond to stimuli through motor movements (Calderon et al., 2016). Arousal levels are measured using both subjective (questionnaires) and objective methods (sensors). Existing bio-sensing technologies can measure pupil dilation, heart rate, electro-dermal activity, brain activity, skin temperature, respiration rate, and other measures of the body's autonomic arousal. IVR literature provides several examples where arousal levels are incorporated into studies on social anxiety (Owens & Beidel, 2015), treatment of phobias (Diemer et al., 2016), presence (Terkildsen & Makransky, 2019), and other studies of emotions and behavior (Marín-Morales et al., 2018; Syrjämäki et al., 2020). However, there are only a few instances in Immersive and non-immersive VR training literature where arousal levels are measured and then linked to performance (Parong & Mayer, 2021; Wu et al., 2010). Such research would open up new avenues for advancing the state of the art, particularly aided by the increasing availability of cost-effective biosensors that can measure physiological arousal and their integration with commercial IVR technologies (e.g., HP Reverb, OpenBCI Galea). If such links can be established, IVR training itself may be further enhanced with adaptation (Zahabi & Abdul

Razak, 2020) by changing the parameters of the training environment to increase or decrease the trainee's arousal levels and performance.

This paper adds to the body of literature on motor skill training in IVR with a between-subjects fine motor skill training experiment. With the aid of N=87 participants, we compared the effectiveness of IVR against physical training conditions with a focus on performance and arousal. The latter is achieved with the use of wearable biosensors which measure physiological arousal in the form of electro-dermal activity (EDA) and electrocardiogram (ECG) signals. These were recorded from all participants across the two conditions. Furthermore, the study investigated improvements in performance after training along with subjective measures of immersion, presence, enjoyment, self-efficacy, and task load.

3.2. Related Works

Training in Virtual Reality

Virtual reality has been described as a collection of technologies that creates synthetic and interactive three-dimensional environments (Mikropoulos & Natsis, 2011). These technologies range from highly immersive ones like head-mounted displays (HMDs) and CAVEs to devices providing a comparatively lower level of immersion like desktops and smartphone displays. Technological advances have resulted in HMDs becoming more popular in recent years, which in turn increased interest in their applications in education and training (Checa & Bustillo, 2020; Makransky & Petersen, 2021). However, research suggests that IVR training should not be just implemented as a one-size-fits-all solution, but instead works best when the design factors of the training environment complement the capabilities provided by the IVR hardware (Jensen & Konradsen, 2018).

Learning/training in immersive virtual environments extends across many domains like school/university education, rehabilitation training for patients, professional training for doctors, and office/industrial workers, where it focuses on diverse kinds of cognitive, affective, and motor skills (Jensen & Konradsen, 2018). For this study, we limit the discussion of training literature focusing on teaching various cognitive and motor skills to healthy individuals. Literature on cognitive skills taught in IVR primarily relates to school and college education (Hamilton et al., 2021), as well as teaching procedural and safety knowledge primarily for industrial training purposes (Feng et al., 2018; Patle et al., 2019). On the other hand, motor skill training literature in IVR has been dominated by medical use cases, particularly in the surgical and dental domains which require fine motor skills (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021). IVR-based motor skill training researchers have investigated the relative advantages IVR-based training has over other training media (physical training, video training, etc.) or variations within IVR, like different levels of visual/haptic fidelity (Huber et al., 2018; Jain et al., 2020), participant characteristics (Shakur et al., 2015), and training methods (Harvey et al., 2019). The results of these studies have been varied; for example, Pulijala et al. (2018) found IVR to be more effective than video/presentation training, Hooper et al. showed IVR to be more effective than physical training for hip arthroplasty surgery, Butt et al. observed the same advantage of IVR over physical training for catheter insertion training, but the advantage disappeared after a week (Butt et al., 2018). Huber et al. found IVR to be as effective as an ‘augmented’ VR condition (Huber et al., 2018). In a comparison of IVR to desktop VR training, Frederiksen et al. found that IVR was inferior in its effectiveness and caused more cognitive load among students of laparoscopic surgery (Frederiksen et al., 2020). Thus, whether IVR training can be as effective or more effective compared to other types of training is inconclusive so far and an open research topic (Checa & Bustillo, 2020) and more so in the case of IVR-based motor skill training (Coban et al., 2022). This need inspired the first research question addressed in this work: *RQ 1 - Is IVR training as effective as physical training in improving task performance?*

In order to answer this research question, it is important to include observable measures signifying training effectiveness (Magill & Anderson, 2016); for example, performance metrics like time for task completion, and quality metrics like the number of mistakes/errors (Abich et al., 2021; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021; Wulf et al., 2010). While measuring such performance metrics, trainees may be tested before and after training to measure their performance improvement (Magill & Anderson, 2016) (p. 269). When the tests are performed in a physical setting, they provide a measure of the transfer of skills from the virtual to the real environment, which has been argued in literature to be crucial in establishing the effectiveness of IVR training (Jensen & Konradsen, 2018; Levac et al., 2019).

Subjective measures have been linked to the effectiveness of learning/training in IVR environments in the Cognitive Affective Model of Immersive Learning (CAMIL) (Makransky & Petersen, 2021). The CAMIL framework suggests that there are two affordances to learning in immersive VR, namely presence (arising from immersion) and agency (arising from interactivity) which affect six other factors, i.e., interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation, which in turn affect the effectiveness of IVR training. Popular subjective measures from IVR training literature, include measures of cognitive load, like the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988), measures of immersion, like the Immersive Tendencies Questionnaire (ITQ), measures of presence, like the presence questionnaire (Witmer & Singer, 1998), measures of usability, like the System Usability Scale (SUS) (Brooke, 1996), measures of cyber/motion sickness, like the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993), and measures of self-efficacy (Lehikko, 2021; Pintrich, 1991). It should be noted that while 'Immersion' is an objective measure of how vivid the VR technology can be made (for example, IVR is more immersive than desktop VR), 'Presence' is understood to be a subjective measure of experience by users which arises from both immersion and interactivity in VR (Makransky & Petersen, 2021). Cognitive load is also crucial to understanding the effectiveness of VR in comparison to other media, as it is negatively correlated with learning/training effectiveness (Koumaditis et al., 2020b; Van Merriënboer & Sweller, 2010). Another subjective measure of importance is 'self-efficacy', defined as the subjective belief people have about their own ability to fulfill a task (Bandura, 1986). Self-efficacy measures are gaining more attention in the literature, as it has been positively linked to the IVR modality and learning outcomes (Shu et al., 2019; Tai et al., 2022). Therefore, it is important to measure the subjective perception of trainees in different training modalities in order to investigate their relationship with training effectiveness. This need generates the second research question: *RQ 2 - Is there a significant difference in the enjoyment, presence, immersion, task load, and changes in self-efficacy reported by participants in IVR compared to physical training?*

IVR training is used in various contexts of motor skills. These can be broadly categorized as context-specific or context-independent. Many examples of context-specific IVR training are found in the medical and surgical domains, where the procedure being trained can easily be used for the same procedure in the real world but rarely in other contexts. An example from the non-medical domain is Winther et al. (2020) who explored the effectiveness of IVR-based training vs conventional training for a pump maintenance task. Such context-specific explorations result in findings that can be applied in the real world easily but are limited by their limited external validity, i.e., they are hard to generalize to other contexts. An advantage of studies on employing context-independent scenarios is therefore that the result is often easier to generalize and transfer to related domains. Examples exist in the IVR motor skill training literature that use more context-independent scenarios like puzzle assembly (Carlson et al., 2015; Koumaditis et al., 2020b; Murcia-Lopez & Steed, 2018). Though such examples are not related to real-world tasks or scenarios, it can be argued that such studies and skill training scenarios may generate results that are more generalizable and transferable to related domains. Inspiration can be found in laparoscopy surgical training literature, where the use of box trainers is widespread, which are highly simplified representations of the tasks involved in laparoscopy (Aggarwal et al., 2004). In this paper, we identify a fine motor skill task (buzz-wire or wire loop game) inspired by literature where it was previously investigated in ergonomics research (Shafti et al., 2016) and in the domain of motor control (Luvizutto et al., 2022; Read et al., 2013) and rehabilitation (Budini et al., 2014; Christou et al., 2018). In this task, the aim is to move a metallic loop across a wire without entering into contact. Immediate feedback is provided when a mistake is made in the form of a loud 'buzz' and, in some cases, a blinking red light in the background. The wire is bent at different locations which makes the task challenging to perform while maintaining a steady hand (Shafti et al., 2016). Read et al. (2013) found that a buzz-wire setup was effective in assessing the relation between manual dexterity and binocular vision. Budini et al. (2014) used buzz-wire training along with hand postural exercises for patients with hand tremors in their experiment and found improvements in goal directed tasks. Christou et al. (2018) presents the only example of research using the buzz-wire setup in an IVR environment, designed as an exercise tool for patients who have suffered stroke and other brain trauma. Similar to Read et al. (2013), they found that the presence of binocular viewing is correlated with increased performance and also that they could distinguish between dominant and non-dominant hand performance. Furthermore, the details provided by Christou et al. (2018) on designing increasing levels of buzz-wire task complexity inspired the current study.

Arousal and Learning

Though the terms 'arousal' and 'emotion' have been used interchangeably in the literature, arousal is one aspect of emotion, along with valence (ranging from negative to positive) according to dimensional models of emotion (Posner et al., 2005; Rubin & Talarico, 2009). Similarly, the terms 'stress' and 'anxiety' have also been used to denote high arousal states with a negative valence (Janelle, 2002; Pakarinen et al., 2019). Multiple methods have been used/utilized to measure arousal levels, using both subjective (Bradley & Lang, 1994) and objective methods (Cacioppo et al., 2007). Among subjective techniques, subjects report their degree of arousal using instruments like the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) and the Stress Arousal Checklist (Mackay et al., 1978). Such questionnaires are usually measured post-exposure and depend on the user's knowledge of their own arousal levels, their memory of the task, and comprehension of the questions. On the other hand, objective measures of arousal are a function of the body's autonomic nervous system, which produces measurable responses, reflecting the user's emotional and cognitive state. This includes changes in skin conductivity (electro-dermal/EDA activity due to sweating), heart rate parameters (heart rate variability/HRV), respiration, skin temperature, pupil dilation, and brain activity (Cacioppo et al., 2007). These bio-signals can be measured by sensors placed on the body (usually non-invasive) to provide measures of physiological arousal. Objective bio-signal data also allow for a more fine-grained look at variations in the subject's arousal levels during a study using measures like Event-Related potentials (ERPs) in EEG, Skin Conductance Responses (SCRs) in EDA, Inter-beat Intervals (or R-R intervals) in Heart Rate Variability data, among many others, where each signal can be used in isolation or be coupled with others in order to increase accuracy (Cacioppo et al. 2007).

Arousal levels may have links to performance and learning outcomes, but limited empirical support is to be found. It has been hypothesized that an individual's experience of arousal affects attention, perception of time, and memory (Storbeck & Clore, 2008), and that there is a non-linear 'inverted U-shaped' relationship between arousal levels and performance (Yerkes & Dodson, 1908). However, the results have been inconclusive in validating this hypothesis (Storbeck & Clore, 2008). Some examples from the literature point to a link between high arousal and better training performance (Homer et al., 2019; Matthews & Margetts, 1991; Ünal et al., 2013). On the other hand, some explorations related to training have found that low arousal leads to better improvements in performance (Kuan et al., 2018; Pavlidis et al., 2019; Prabhu et al., 2010; Quick et al., 2017). The link between arousal and learning/training adds a further layer of complexity since the effectiveness of training is measured not by task performance alone but by changes in performance across different periods, usually as a change in performance before and after training (learning gain). Movahedi et al. (2007) illustrate this complexity in a sports training context where they found that participants performed worse during a retention test when their arousal levels during the test were mismatched with the arousal levels (either high or low) during training.

The use of physiological data to measure arousal levels in IVR literature is rare; however, some representative examples that use heart rate-related metrics for measuring arousal include Muñoz et al. (2019) where HRV metrics (along with EEG data) were used to detect calmness states among participants using an IVR target shooting simulator, Cebeci et al. (2019) where eye tracking and heart rate were used to measure the impact of different virtual environments on factors like cybersickness and emotions among study participants, and Larmuseau et al. (2020) where HRV along with EDA and skin temperature were used to measure cognitive load among students' learning statistics online. In the use of EDA data, some illustrative examples include understanding how soldiers respond to threatening stimuli during IVR training (Binsch et al., 2021), detecting student stress levels during a physics course (non-VR) (Pijeira-Díaz et al., 2018), and measuring EDA responses to insights made by participants in an IVR learning environment (Collins et al., 2019). There are currently only a few examples in IVR literature on the exploration of physiological arousal levels and their connection to fine motor skill training in virtual reality. One example is from a science education scenario where it was shown that learning in IVR leads to higher arousal and subsequently lower scores on a retention test (Parong & Mayer, 2021). Another example is from non-immersive VR where a Stroop interference task induced arousal in participants during a virtual driving task and then found the optimal arousal levels related to increased performance (Wu et al., 2010). Therefore, a research gap exists in the literature for understanding the link between motor skill training in IVR, improvements in performance due to the training, and physiological arousal levels of the trainees. The following research questions were generated in order to address this gap: *RQ 3 – Is there a significant difference between the physiological arousal levels of participants in IVR training compared to physical training? RQ 4 – Is there a link between physiological arousal during training and improvements in performance after training?* In the next section, the design of the experiment is detailed which will help address these questions.

3.3. Methods

The experiment contains three phases as depicted in Fig. 33: a pre-training phase common to all conditions where a pre-test of the motor skill is performed, a training phase in which the participants were randomly assigned to either VR or physical training conditions and a post-training phase where a post-test of the motor skill was performed for participants from both training conditions. The following sub-sections details the motor skill task, the two experimental conditions, the pre-test and post-test tasks, the physiological and performance data measured during the experiment as well as the subjective data reported by the participants. The section ends with a detailed description of the experimental procedure shown in Fig.1.

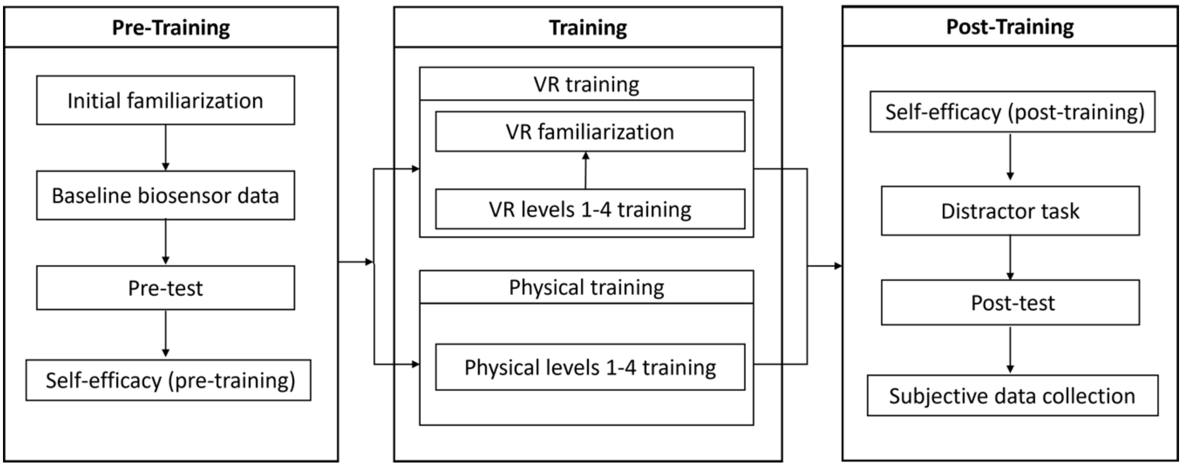


Figure 33. Overview of experiment procedure.

3.3.1 Motor skill task

In this study, the trainee is asked to grab the apparatus as seen in Fig. 34 and guide the metallic loop across a wire as fast as possible with the least amount of touching between the loop and the wire. There are two variations of the task, varying on the feedback provided when the loop touches the wire, i.e., when a mistake is made. In the training task, when the participant makes such a mistake, three kinds of feedback were provided simultaneously:

- *Haptic feedback* in the form of vibration in the Oculus Quest's Touch controller. Vibration is set to the maximum frequency and amplitude available in the Oculus SDK and delivered for 1/10th of a second.
- *Auditory feedback* was provided by playing a continuous 1000 Hz sine wave tone at 39 dB over the headphones worn by the participant (Sony WH-CH710N). Sound levels were verified and maintained across participants using the NIOSH iPhone app (National Institute for Occupational Safety and Health Sound Level Meter App).
- *Visual feedback* is provided by switching on a red LED (Fig. 34) placed at eye level behind the wire.

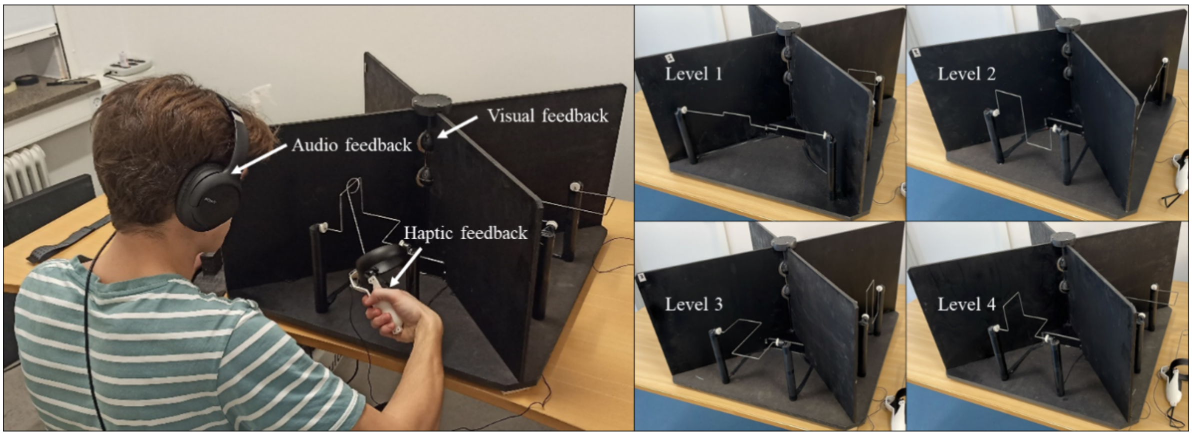
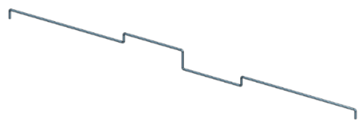
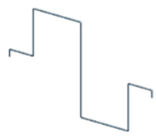
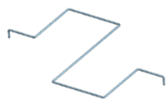
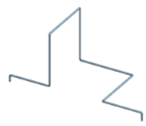


Figure 34. Physical training condition. Left: participant moving loop across the wire in level 4. When the loop touches the wire, the participant receives audio, haptic, and visual feedback. Right: The four levels of training.

Table 12. Training levels

Difficulty Level	Movement Pattern	Level Design
1	The first level (48 cm long) is almost straight across the x-axis with short deviations in the y-axis. The participant can complete the task with minimal twisting of the wrist.	
2	The second level is 52 cm long and has bends in the y-axis. Participants may have to twist their wrists substantially compared to level 1.	
3	The third level is 52 cm long (similar in proportions to level 2) with bends in the z-axis.	
4	The last and most challenging level is 48 cm long with bends on all three axes.	

The training task in the physical and VR conditions is spread across four levels of increasing difficulty, with difficulty being specified as an increase in complexity of wrist movements needed to complete a level (see Table 12). For example, a wire with fewer bends requires less wrist movement, which in turn may produce fewer mistakes (i.e., the loop touching the wire) and the task may be completed (move from start to finish) quicker than a wire with more bends. This was verified in a previously published pilot study (Unnikrishnan Radhakrishnan, Alin

Blindu, et al., 2021). These four levels were intended to help the participants train themselves, i.e., to develop the skills required to perform the test task more effectively. It should be noted that there were no instructions provided in either condition to facilitate the training by letting the participant construct their strategies for improving their skill level subject to the constraint of the environment.

3.3.1.1 Training in physical condition

The wire in each training level rests on two 20cm tall pillars to provide better task ergonomics for participants (verified in a pilot test). Two black vertical wooden panels are placed at right angles on the wooden base (Fig. 34) and the entire setup is painted black to reduce visual distractions. The start and end positions are shaped like cylinders with grooves inside for the loop to be placed. An Arduino Uno placed in a microcontroller box is used to detect contact between the loop and the wire (denoting mistakes) using a simple switch circuit. A “mistake” signal is transmitted serially to the PC when the loop touches the wire. Two similar switch circuits are used to detect contact between the loop and the grooves on both the start and finish positions. When the participant lifts the loop off the start position, a “start task” signal is transmitted by the contact circuit to the PC; similarly, an “end task” signal is transmitted when the loop is placed in the end position. The loop is made by bending a 1mm thick metal wire with a diameter of 2.5cm. The loop is then screwed to a 3D printed handle (adapted from Lagos (2019)) that houses an Oculus controller (Fig. 34) to provide haptic feedback.

3.3.1.2 Training in IVR condition

Participants in the VR condition wore an Oculus Quest (1st generation) Head Mounted Display (HMD) (Fig. 35) connected to a PC and running on Rift mode. The VR environment was developed in the Unity3D (version 2019.4) game engine to closely resemble the physical environment. The wires (for each training level) and loop were designed using the Blender3D design software. The participants were presented with the same four levels in VR as in physical condition. They hold a physical handle containing the loop and the Oculus controller (like those in the test and physical training tasks). The position and rotation of both the controller and the HMD are provided by the Oculus SDK which is then used to move the virtual loop and the participant’s viewpoint in the three-dimensional space of the virtual environment (see Fig. 35).

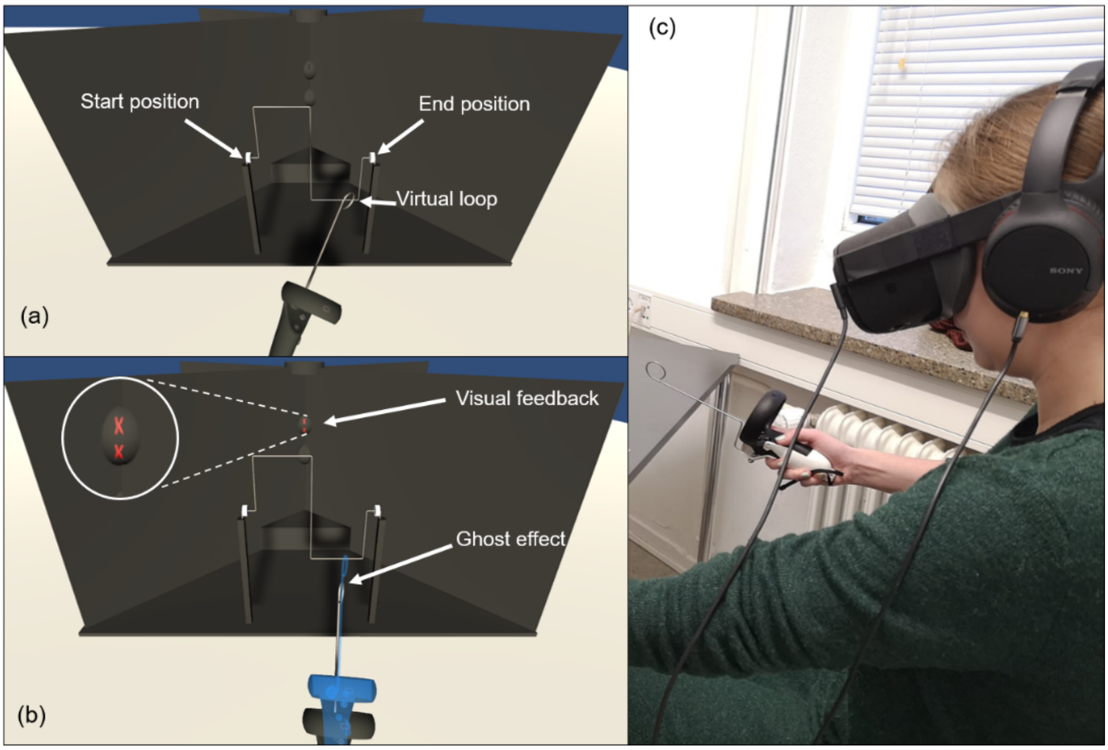


Figure 35. (a) VR training environment. Virtual loop moving across level 2, (b) Ghost loop appears when contact is made, and the "real" loop goes outside the wire. It disappears when the loop is placed back inside the wire. Visual feedback in form of a red 'X' mark in the background also turns on during contact. (c) Participant in VR condition wearing an Oculus Quest HMD (Rift mode).

The "Measurements" asset from the Unity Asset Store was used to scale and position objects identically to their real-world counterparts (vrchewal, 2020). Both haptic and audio feedback modalities used the same parameters as the physical condition, and the visual feedback was in the form of a red 3D light behind each wire turning on during contact between the virtual wire and the virtual loop (Fig. 35-b). Like the physical condition, "start task," "end task," and "mistake" signals were sent to the data collection module (Fig. 37). Physics collision meshes were defined on the 3D models of the loop, the start and end positions, and the wires across the four levels. Collision tests were performed by Unity's inbuilt physics engine at 60 Hz.

Though the VR condition mimics the physical, there are unavoidable differences between the two conditions:

- *Ghost effect during mistakes* – When the participant makes a mistake, i.e., the loop touches the wire, there is nothing to physically restrict the participant's hand, unlike the physical condition where there is an actual wire to provide resistance. Though there is haptic vibration when contact is made, by the time the mistake is made, the loop would have passed through the wire creating an unrealistic effect for the participant which could potentially break their feeling of immersion (i.e., "being there"). To solve this, a "ghost effect" has been programmed to show a blue translucent loop at the contact position where the actual loop passes through the virtual wire (Fig. 35-b). This helps the participant understand how to bring their loop back into the wire, at which point the blue translucent 'ghost' disappears.

- *VR familiarization* – Participants were first exposed to a VR task to help them familiarize themselves with the movement of the virtual loop before starting the actual training. This is to avoid any negative outcomes from the novelty effect of using IVR among novice users (Hamilton et al., 2021). They were encouraged to intentionally make mistakes to learn the functionality of the ghost effect. The task is in the form of a straight wire which has no bends so that there is no unintended extra "training effect" for participants in the VR condition.
- *Differences in media* – In addition to the above two features which distinguishes VR from the physical, there are other differences arising from the nature of the VR medium itself, for example - the field of view and the visual acuity provided by the Quest HMD is lower compared to that provided by healthy human vision (Adhanom et al., 2021; Cuervo et al., 2018). Additionally, the weight of the HMD has not been replicated in the physical condition.

3.3.1.3 Test task

The wire in the test task is 52 cm long with eleven 90-degree bends in all three axes (x, y, and z) between the start and finish positions (see Fig. 36). Contact circuits like those used in the physical training setup are used here to detect contact between the loop and the wire as well as the corresponding start and end positions. The three contact signals "start task", "mistake", and "end task" are serially transmitted to the PC similar to the training setup (see 3.1.2). There is no "augmented" feedback provided when the participant makes a mistake in the test condition, i.e., there is no haptic, visual, or auditory feedback other than the natural feedback of two metal pieces touching each other. Like the training setup, all parts of the test setup are painted black to provide a consistent background with fewer visual distractions. An Oculus controller is placed inside the handle containing the loop to mimic the weight of the controller in the physical and VR training setups but provides no haptic feedback. The same test task is used before and after the VR/physical training task as an objective measure of training effectiveness.

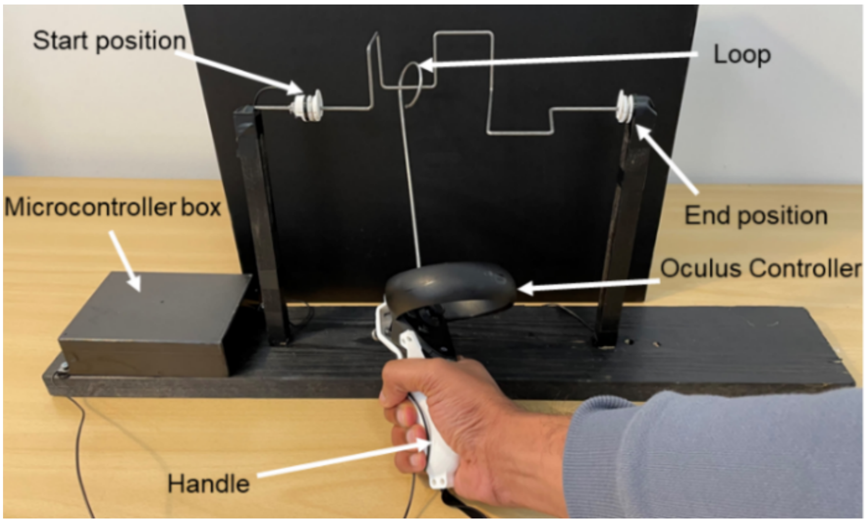


Figure 36. Test task setup along with the loop attached to 3D printed handle containing a Quest controller.

3.3.3 Sensors and data collection

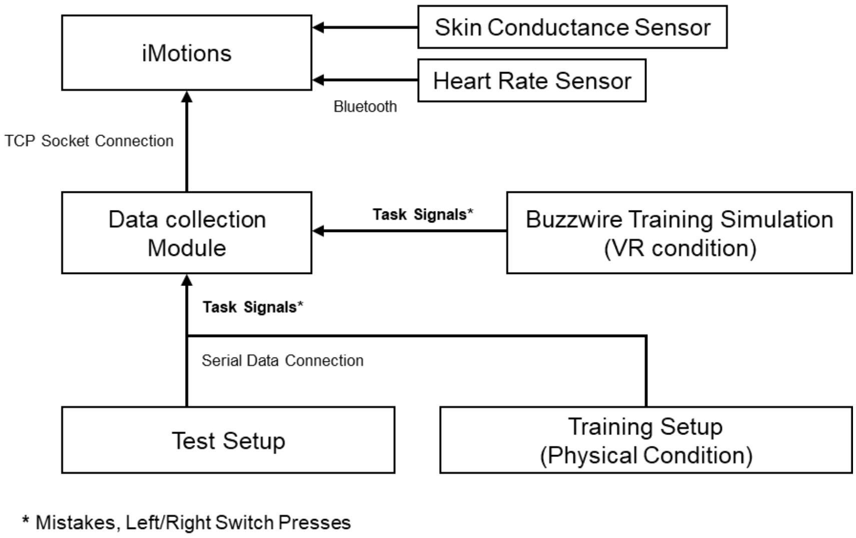


Figure 37. Software Architecture

Data collected during the experiment comes from three kinds of sources: the biosensors, the task-related signals coming from the test and training setups, and subjective data recorded in an online survey (at the end of the experiment). The first two types of data are facilitated by:

- *iMotions* – iMotions is a commercial software platform that supports data collection from commercial biosensors across many modalities (iMotions A/S, Copenhagen, Denmark). In this study, iMotions was used as the endpoint for storing all data coming through the dataflow pipeline shown in Fig. 37, as it integrates timestamped data from the two biosensors alongside performance-related data coming from the data collection module.

- *Data collection module* – A data collection module was developed in C# on the Unity3D game engine which collected task-related signals from the hardware setups (test and training) and the VR training software. Data from the hardware was read from two serial connections with a transmission rate of 9600 baud. The data collection module then transmitted in real-time the collected signals to the iMotions biosensor platform via a TCP socket connection (Fig. 37).

Subsequent subsections discuss the biosensors used for measuring electro-dermal and heart rate signals, associated arousal metrics (3.3.1), performance metrics for measuring the effectiveness of training (3.3.2), and survey data to measure the subjective experience of training using online questionnaires (3.3.3).

3.3.3.1 Physiological sensing

For measuring the participant's physiological arousal levels, the Polar H10 (heart rate) and the Shimmer GSR+ (skin conductance) sensors were used. Table 21 in the appendix details all the physiological metrics used, their source, and their relationship with arousal according to literature.

Electrocardiogram (ECG) signals

The Polar H10 (Polar Electro Oy, Kempele, Finland) is an Electrocardiogram (ECG) based Heart Rate (HR) monitor designed for athletes. It has been clinically validated to be as effective as medical-grade ECG hardware (Gilgen-Ammann et al., 2019) and has been used in recent VR literature (Muñoz et al., 2019; Ventura et al., 2021). It is worn around the chest with electrodes placed in contact with the skin. The data in the form of heart rate and Inter-beat Intervals (R-R intervals) are transmitted via a Bluetooth Low Energy (BLE) connection at a rate of 1-2 Hz to the iMotions application running on a PC. Measures of heart rate variability including time and frequency domain metrics have been calculated using the hrv-analysis Python library (Champseix 2022).

Increases in arousal are indicated by increases in Heart Rate (time-domain) and frequency-domain measures like LF/HF (Low Frequency/High Frequency) Ratio (Orsila et al., 2008; Slater et al., 2006). On the other hand, decreases in time-domain HRV measures like IBI (Inter-Beat Interval), SDNN (Standard Deviation of NN Intervals), RMSSD (Root Mean Square of Successive Difference), and the frequency domain measure HFN (Normalized High-Frequency

Component), indicate an increase in arousal (Shaffer & Ginsberg, 2017). All HRV metrics have been baseline corrected by subtracting from them the corresponding mean baseline values (Healey & Picard, 2005; Wulfert et al., 2005).

Electro-dermal activity (EDA) signals

The Shimmer GSR+ (Galvanic Skin Resistance) unit (Shimmer Research Ltd., Dublin, Ireland) measures EDA by passing a small current through electrodes placed in two locations on the body. The locations for the electrodes were verified in a pilot study where Shimmer electrodes were placed on the foot, the forehead, and the fingers of two participants, and the signals generated in response to stimuli were examined for signal quality and consistency. It was found that the index and middle fingers were the most reliable locations for sensing skin conductance which matched recommendations from literature on skin conductance sensing (van Dooren & Janssen, 2012). The index and middle fingers of the left hand were chosen to allow study participants to use their right hand alone for moving the loop across the wires.

Popular EDA measures include SC (Skin Conductance) measured in micro-siemens which increases in response to an increase in arousal (Collet et al., 2005). An increase in arousal also leads to a higher rate of skin conductance responses which are peaks in the SC amplitude lasting between 1-5 seconds after onset (Krogmeier et al., 2019; Terkildsen & Makransky, 2019). Accordingly, the SCRPeaks metric is calculated as the number of skin conductance response peaks per minute. Similarly, the mean peak amplitude of all SCR peaks (SCRAmp) is also a positive measure of arousal (Khalfa et al., 2002; Krogmeier et al., 2019). SCL levels have been baseline corrected by subtracting from it the mean baseline values (Potter & Bolls, 2012). All EDA signals were processed using the Neurokit2 Python library (Makowski et al., 2021).

3.3.3.2 Improvement in performance

The data collection module collects signals generated from both physical and IVR setups, namely the “Start task”, “End task”, and “Mistake” signals. These are used to calculate the following two measures of performance:

- *Task completion time (TCT)* – The time taken to move the loop from start to end.
- *Contact time (CT)* – The total time the loop is in contact with the wire during the task which quantifies the number of mistakes by the participant.

These two measures are then used to calculate the following measures of performance improvement:

- *Improvement in task completion time (TCT-I)* – This is calculated by subtracting the posttest TCT from the pretest TCT for each participant. A positive value indicates an improvement in this performance metric.
- *Improvement in contact time (CT-I)* – This is calculated by subtracting the posttest CT from the pretest CT for each participant. A positive value indicates an improvement in this performance metric.
- *Improvement Score (IS)* – Since the participants are asked to complete the test task by satisfying two potentially competing goals – to minimize both task completion time and contact time – participants may choose to prioritize one over the other. For example, a participant can choose to complete the task very slowly to minimize the chances of contact with the wire or vice versa. To balance out these two metrics, it is necessary to create a combined score metric that considers both improvements in task completion time (TCT-I) and contact time (CT-I). To calculate this measure, we first divide the two performance improvement measures, TCT-I and CT-I, into 10 equal-sized quartiles for all participants across both conditions, transforming the values into scores from 1 to 10 where 1 denotes the least improvement in performance and 10 the most. Subsequently, IS for a participant is defined as the sum of these two scores. A hypothetical participant who has improved the most in both TCT-I (score = 10) and CT-I (score = 10) metrics would then get a final improvement score (IS) of 20.

3.3.3.3 Subjective data

Subjective data was collected from all participants towards the end of the experiment using an online survey tool (Microsoft Forms) running on a lab PC. The different subjective metrics are listed below.

- *NASA Task Load Index (NASA-TLX)* – NASA Task Load Index (Hart & Staveland, 1988) is a validated measure of workload across six dimensions (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration). The 'raw' version of the NASA-TLX without weighted rankings was given to the participants where the answer to each measure was on a scale of range 1-21 (Hart, 2006).

- *Immersion Questionnaire* – The immersion questionnaire from Högberg et al. (2019) was adapted. Participants are asked to give answers on a Likert scale ranging from 1 to 7 (from strongly disagree to strongly agree). A combined Immersion Score is calculated by taking the average of all the responses to items. See Appendix for a list of all items in the questionnaire.
- *Presence Questionnaire* – The presence questionnaire was adapted from the physical presence subscale of the Multimodal Presence Scale (Makransky et al., 2017) and the telepresence questionnaire (Kim & Biocca, 1997). Participants are asked to give answers on a Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). A combined Presence Score is calculated by taking the average of all the responses to items (after reversing responses to inverse questions). See Appendix for a list of all items in the questionnaire.
- *Enjoyment* – The participants are asked to rate their agreement with the question *"The training session was very enjoyable"* on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).
- *Self-efficacy* – The participants are asked *"How confident are you that you can perform a similar task effectively (go from start to finish as fast as you can with minimal mistakes) on a scale from 1 to 7?"* to measure self-efficacy, once before the training and once after training. Details are provided in the next section.

3.3.4 Study procedure

Ethics approval was obtained from the local research ethics committee for experimenting with human subjects. The study was conducted in two rooms, one dedicated to IVR training and the other to physical training. Participants signed up for the study using the lab's online participant recruitment system. The system automatically filtered the participants using the following criteria based on self-reported data (i.e., they were not medically certified or independently verified): (a) right-handed, (b) normal vision or corrected to normal vision with contact lenses, and c) no mental illnesses or sensitivity to nausea. The requirement for right-handedness was added to eliminate variation in the setup. Participants signed up for 45-minute timeslots of their choosing and were paid the equivalent of 15 Euros. Each condition/room was run by one researcher at a time. The researchers switched between them regularly to reduce investigator effects. The timeslots for both conditions were open from 9 a.m. to 5 p.m. on weekdays.

At the beginning of a session, the participants were asked to read and sign the consent form. They were then briefly familiarized with the experiment procedure by allowing them to practice on the first level of the physical training setup. Thus, all participants, independent of condition, were provided a chance to experience the physical setup (Fig. 34), the cue for starting each task (when they hear the word 'Go'), and the proper way to lift the handle from the start position and to rest it on the end position. Thereafter, they were given the privacy to wear the Polar H10 around their chest as the researchers left the room. After this, the Shimmer GSR electrodes were placed on the index and middle fingers of the participant's left hand. The participant places her/his left hand on a Styrofoam support pad placed towards the left side of the table with the palms facing upwards and the fingers kept relaxed. The participant was asked not to move or flex her/his hand to minimize the noise in the recorded signals. The signal quality for both sensors was checked and verified in the iMotions software before the experiment started.

Baseline biosensor data was then measured by asking the participants to remain seated quietly and still with their eyes closed, without heavy breathing. The baseline HR and GSR data were then used to normalize subsequent signals since the baseline HR and GSR values for each person varied considerably. The participants were then presented with the test task before training begins (detailed in section 3.1.1.). They start the test task after hearing the word 'Go' from the researcher. Upon completion, they were then asked the question on self-efficacy. Following this, they were trained on four levels of increasing complexity in either VR or physical conditions (depending on the random assignment at the beginning of the experiment). In the physical condition, after each level of training, the researcher would rotate the wooden base by 90 degrees (Fig. 34) so that the next level is facing the participant. This process took 10 to 15 seconds, which was absent in the VR condition where the switch to the next level was instantaneous. At the beginning of each level, they were asked to relax for 30 seconds by resting their right hand on their lap and start the task only when they hear the word 'Go', this time from the headphone. After the training, the participant was asked the self-efficacy question again. They were then presented with a distractor task in the form of a maze to reduce the recency effect (Carlson et al., 2015; Winther et al., 2020). They were asked to spend about a minute both visualizing the solution and then picking up the maze with their right hand to solve it, in order to minimize recency effects (Bjork & Whitten, 1974). They were finally given the test task and again asked to perform it as quickly as possible with the least number of mistakes possible. Following this, the participant was asked to remove the sensors and to fill out an online questionnaire containing the NASA-TLX questionnaire and questions on enjoyment, presence, and immersion. When the participant started performing either the test or training task, the researcher steps behind a panel to reduce biases in performance due to the Hawthorne effect (Demetriou et al., 2019).

No personal information was recorded, except for those required for compensating the study participants, which were handled according to university data protection policies. The researchers followed Covid-19 safety protocols, including sanitizing the sensors, table, and buzz-wire handles after every participant completed the experiment.

3.4. Results

The statistical analysis was performed using the statistical methods available in SciPy (Scientific Python) and Pingouin packages (Vallat, 2018; Virtanen et al., 2020), and plots were generated using the Seaborn and Matplotlib Python packages (Hunter, 2007; Waskom, 2021). 87 participants were part of the study, divided between the physical (N=42) and VR training (N=45) conditions. 48 participants identified themselves as male, 37 as female, and 2 as other. 46 participants indicated their age group in the 18-24 range, and 36 indicated theirs in the range 25-34. The majority of participants in the VR condition (69%, N=31) indicated that they had tried a VR head-mounted display 1-5 times, 1 reported trying IVR 5-10 times, and 6 reported trying IVR more than 10 times, whereas 7 had never tried VR before. Data from eight participants had to be excluded from the analysis of performance metrics because of data loss arising from VR headset tracking errors, and the biosignals from 15 participants had to be excluded from analysis due to sensor errors. Shapiro-Wilk tests for normality were applied to all the variables, and if a variable was found to violate assumptions of normality, non-parametric statistical tests were used: Wilcoxon Signed Rank (Wilcoxon, 1945) for paired, and Mann Whitney U tests for independent tests (Mann & Whitney, 1947), and the related W and U statistics are reported. When the variables used for comparison, both followed normal distributions, Student's t-test and Welch's t-test (for unequal variances) were used to test for independence, and related t-statistic and Cohen's d are reported. A significance level of 0.05 was selected while interpreting the results of the statistical tests. The datasets analyzed during the current study are available from the corresponding author on reasonable request.

3.4.1 Improvement in performance

Fig. 38 depicts the three performance metrics for the VR and physical conditions: task completion time (Fig. 38a), contact time (Fig. 38b), and improvement score (Fig. 38c) (see section 3.3.3.2 for definitions). The task completion time and contact time metrics were analysed to see if there were changes from the pre-training task to the post-training task. Analyses were also performed to see if there were statistically significant differences between the improvement scores of the two conditions.

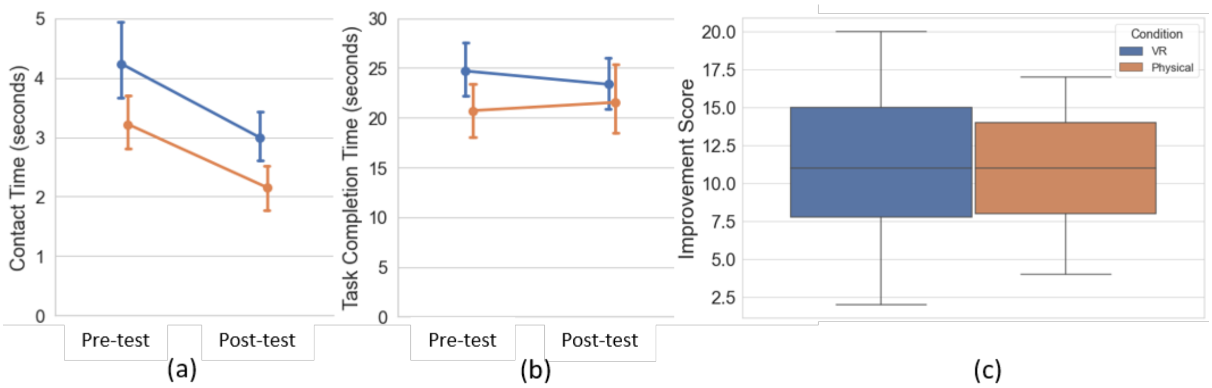


Figure 38. Change in performance metrics within VR (N=45) and physical conditions (N=42) for (a) contact time, (b) task completion time, and (c) between the conditions for improvement score.

Within-condition changes

In terms of contact time (CT), a statistically significant decrease of 1.21s from pre- to post-training ($p<0.001$, $w=126.0$) was observed among participants in the VR condition (N=40). For the same group, a near statistically significant decrease of 1.33s was observed in the task completion time (TCT) from pre-training to post-training phases ($p=0.062$, $w=352.0$). In the physical condition (N=39), there was a statistically significant decrease of 1.07s in CT from pre- to post-training phases ($p<0.001$, $w=114.0$). On the other hand, though a slight deterioration of TCT may be observed in Fig.38-b for the physical condition from pre-training to post-training phases, this was not statistically significant (0.83s, $p=0.412$, $w=387.0$).

Between conditions

To compare performance in VR (N=40) and physical (N=39) conditions, improvements in task completion time (TCT-I), contact time (CT-I), and improvement scores (IS) were calculated (see section 3.3.2). Since the metrics from both these conditions were non-normally distributed, Mann-Whitney U independent samples tests were performed. The results showed no statistically significant differences between the improvement scores in the two conditions ($p=0.353$, $t(77) = -0.38$, $d = 0.085$). Regarding improvement in task completion time (TCT-I), though it can be seen from Fig.38-b that the task completion time for participants in the VR condition shows a visible improvement (i.e. decreases), this was not statistically significant ($p=0.2864$, $U = 722$). CT-I also showed similar trends with participants in the VR condition showing no statistically significant differences with participants from the physical condition ($p=0.4746$, $U = 773$).

3.4.2 Improvement in self-efficacy

As indicated in Fig. 39, in the VR condition (N=45), there is a statistically significant increase in the reported self-efficacy from the pre-training phase (3.8) to the post-training phase (4.24; $p=0.016$, $w=120.5$). Though a slight increase in reported self-efficacy in the physical condition (N=42) from the pre-training phase (4.38) to the post-training phase (4.48) can be observed in Fig.39, this difference was found not to be statistically significant ($p = 0.545$, $w = 191.5$). It was also observed that the change in self-efficacy in the VR condition (0.44) was greater than the change in self-efficacy in the physical condition (0.095). This difference approaches statistical significance ($p = 0.0585$, $U = 767.5$).

3.4.3 Task load

Fig. 40 shows the item-wise scores for NASA-TLX between the VR (N=45) and physical (N=42) conditions. Participants reported their perceived task load on six dimensions, i.e., mental, physical, and temporal demand, along with frustration, effort, and performance (Hart & Staveland, 1988). Among these six dimensions, it can be observed that both VR and physical training result in similar task load values except for the temporal load parameter where participants in the physical condition report a mean score of 11.71 ± 2.87 (on a scale from 1 to 21) which is significantly higher than what participants in the VR condition reported (9.16 ± 2.49; $p=0.012$, $U = 738.5$). There was no statistically significant difference in the combined NASA TLX Score between the physical (11.62 ± 2.87) and VR conditions (11.53 ± 2.49; $p=0.436$, $t(81.5) = 0.161$, $d = 0.03$).

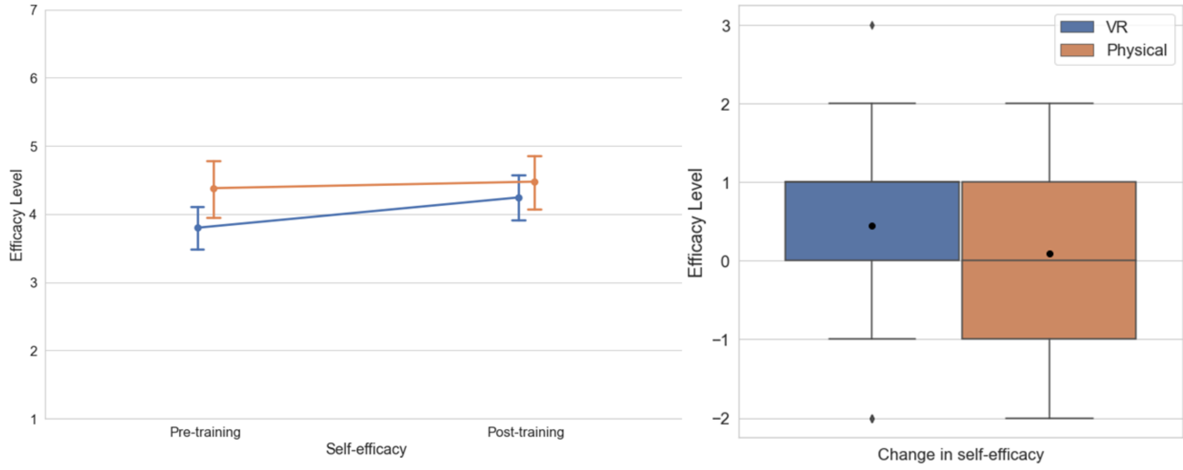


Figure 39. Left: Self-efficacy levels from pre-training to post-training phases (on a scale of 1-7). Right: Change in self-efficacy levels across VR (N=45) and physical conditions (N=42).

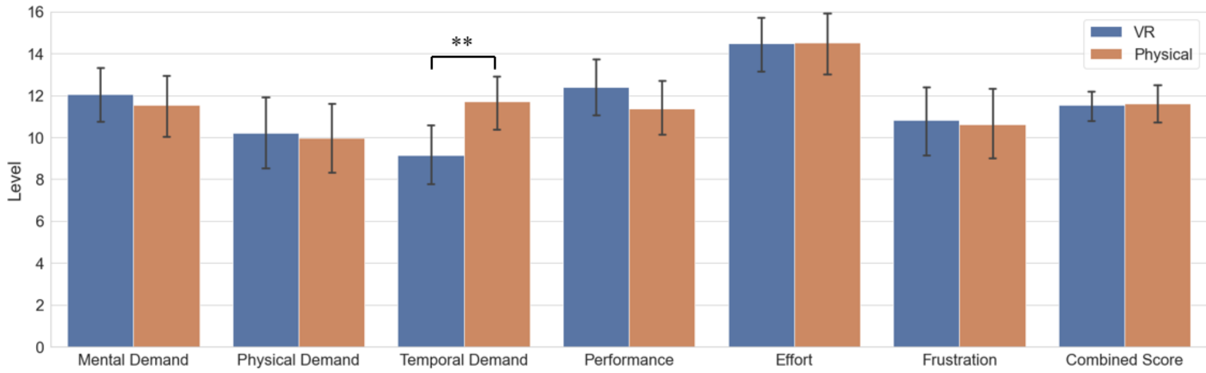


*Figure 40. NASA TLX Scores across VR (N=45) and physical conditions (N=42). ** denotes significant difference at α = 0.05*

3.4.4 Immersion, presence and enjoyment

Fig. 41 shows the Immersion, Presence, and Enjoyment scores between the VR (N=45) and physical (N=42) conditions. Cronbach's alpha coefficients were calculated for both questionnaires and found to be 0.88 for Immersion and 0.69 for Presence, indicating an acceptable internal consistency of the scales. An analysis of the Immersion Score (which is the mean of all items on the Immersion questionnaire) shows that participants in the VR condition report higher immersion on average (4.94 ± 0.99) as compared to participants in the physical condition (4.54 ± 0.98) and that this difference is statistically significant (p=0.031, t(84.47) = -1.88, d = 0.404) with statistical significance also being observed for items I2, I4, and I9. Analysis of the combined Presence Score shows participants reporting a higher score on average for VR (4.61 ± 0.93) compared to physical (4.4 ± 0.79). This difference approaches statistical significance (p=0.0736, U = 774) with statistical significance also being observed for items P5, P6, P10, and P14. See Tables 25 and 26 in the Appendix for item-wise statistics for both Immersion and Presence questionnaires. Finally, participants report higher enjoyment for the VR condition (6.02 ± 1.23) as compared to physical condition (5.52 ± 1.15; p=0.0175, U = 696.5).

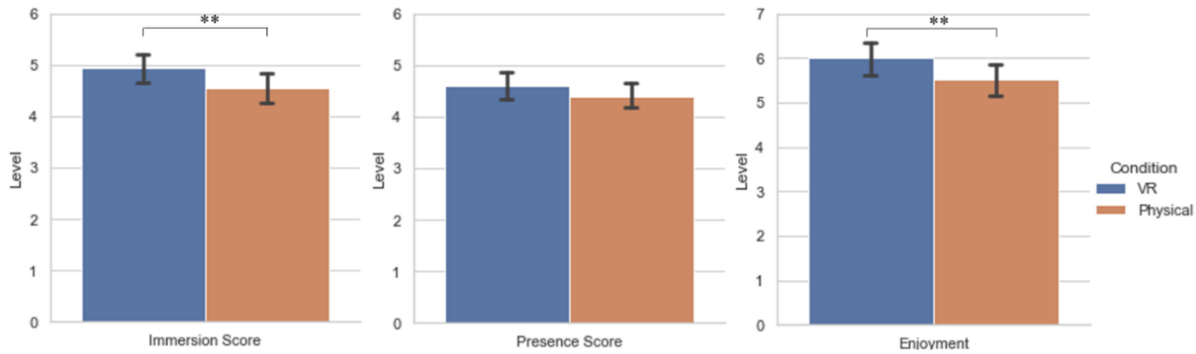


*Figure 41. Immersion, presence, and enjoyment scores across VR (N=45) and physical conditions (N=42). ** denotes significant difference at α = 0.05*

3.4.5 Physiological arousal

Arousal levels between conditions

Table 13 lists all the physiological arousal metrics recorded during the training session. Only data points recorded between the start and finish points for each training level have been considered and then averaged to generate arousal metrics that represent the whole training phase. The metrics listed have been adjusted to each participant's baseline where appropriate.

*Table 13. Physiological arousal metrics across physical (N = 39) and VR training conditions (N = 39 for HRV, N=33 for EDA). ** denotes significant difference at α = 0.05, # denotes baseline corrected metrics.*

Physiological Arousal Metric		Physical (Mean ± SD)	VR (Mean ± SD)	p-value
HRV	Mean HR#	0.05±3.6	-1.31 ± 4.29	0.066
	Mean IBI#	1.33 ± 48.15	16.86 ± 51.73	0.087
	Mean RMSSD#	-8.6 ± 172.06	-22.2 ± 127.24	0.614
	Mean SDNN#	-9.0 ± 125.07	-18.34 ± 79.84	0.3821
	Mean LF/HF Ratio#	-1.15 ± 3.52	-1.02 ± 5.0	0.3086
	Mean HF Normalized#	9.72 ± 17.32	5.83 ± 22.19	0.1737
EDA	Mean SC#	2.63 ± 2.11	2.31 ± 2.21	0.1222
	Mean SCRAmp	0.21 ± 0.21	0.19 ± 0.16	0.445
	Mean SCRPeaks	12.04 ± 3.31	9.9 ± 2.29	**0.0032****

Among EDA metrics, in the VR condition (N=33), the mean SCRPeaks of 9.9 was found to be significantly lower than the SCRPeaks of 12.04 in the physical condition (N=39) denoting higher arousal among participants in the physical condition ($p = 0.0032$, $U = 885$). Among HRV measures, mean baseline-corrected HR was lower in VR (-1.3) than physical (0.05) and the difference approaches statistical significance ($p = 0.066$, $t(73.7) = 1.52$, $d = 0.34$). Showing similar trends, the mean baseline-corrected IBI was found to be higher in VR (16.86) than physical (1.33) but the difference is not statistically significant ($p = 0.087$, $t(75.6) = -1.37$, $d = 0.31$).

Comparisons between other EDA and HRV metrics showed no statistically significant differences though they mostly align with the findings in the SCRPeaks and IBI metrics with higher arousal in physical than in VR. Among EDA measures, the mean SC across all training levels in the VR condition (N=33) is 2.31, which is lower than the mean SC from the physical condition (N=39), 2.63. However, this difference is not statistically significant ($p = 0.1221$, $U = 747$). Mean SCRAmp for VR (0.19) is lower than physical (0.21) but the difference is not statistically significant ($p = 0.445$, $U = 676$). Among time-domain HRV measures, the mean baseline-corrected RMSSD for VR (-22.22) is higher than physical (-8.6) with no statistically significant difference ($p = 0.614$, $U = 732$) and the mean baseline-corrected SDNN in VR (-18.34) is lower than physical (-9.0) with no statistically significant difference ($p = 0.3821$, $U = 791$). Among frequency domain HRV metrics, mean baseline-corrected HFN in VR (5.83) is lower than physical (9.72) where the difference is not statistically significant ($p = 0.195$, $t(71.8) = 0.865$, $d = 0.196$) and mean baseline-corrected LF/HF Ratio in VR (-1.02) is greater than physical with no statistically significant difference ($p = 0.3086$, $U = 710$).

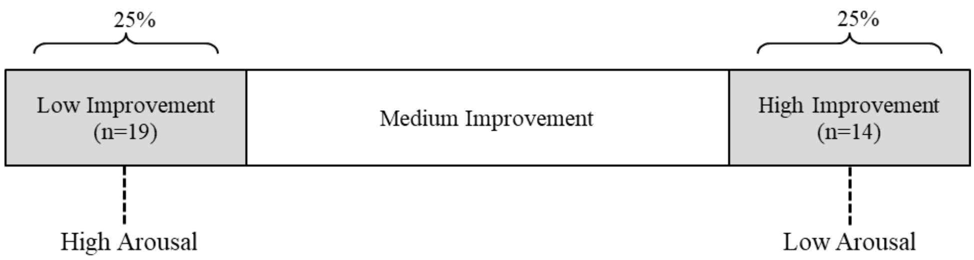


Figure 42. The participants (from both conditions) were divided into high and low-performance groups. The high improvement group is in the upper 75th percentile of performance based on the improvement score. Similarly, the low improvement group is from the bottom 25th percentile. Participants who showed the highest improvement had lower arousal than those who had the lowest improvement.

Arousal level and performance

To assess the link between arousal levels and performance, data from both IVR and physical groups were combined, and Spearman rank correlation tests (for non-normal data) were performed between the physiological arousal metrics and performance improvement metrics. The tests showed almost no correlation between arousal and improvement in performance with most ρ values between -0.1 and 0.1. Notable statistically significant but weak correlations include the correlation between TCT-I and SCRAmp (ρ = -0.24, p = 0.041), TCT-I and SC (ρ = -0.24, p = 0.0434), and near statistically significant correlations include those between TCT-I and RMSSD (ρ = -0.21, p = 0.068) and IS and SCRAmp (ρ = -0.19, p = 0.098).

*Table 14. Physiological Arousal Metrics across High (N=14 for HRV, N=11 for EDA) and Low Improvement groups (N=19 for HRV, N=18 for EDA). ** denotes significant difference at α = 0.05, # denotes baseline corrected metrics.*

Physiological Arousal Metric		High Improvement (N=14) (Mean ± SD)	Low Improvement (N=19) (Mean ± SD)	p-value
HRV	Mean HR#	-0.11 ± 3.17	-1.24 ± 3.76	0.196
	Mean IBI#	6.17 ± 46.41	14.11 ± 53.97	0.196
	Mean RMSSD#	38.69 ± 274.99	-57.28 ± 120.41	0.6215
	Mean SDNN#	27.15 ± 201.54	-38.22 ± 73.97	0.5935
	Mean LF/HF Ratio#	-2.02 ± 7.78	-0.37 ± 1.41	0.7013
	Mean HF Normalized#	0.58 ± 22.21	7.15 ± 19.15	0.2167
EDA	Mean SC#	1.59 ± 0.9	3.49 ± 2.74	**0.0252****
	Mean SCRAmp	0.12 ± 0.09	0.25 ± 0.18	**0.0298****
	Mean SCRPeaks	12.03 ± 2.29	11.92 ± 2.85	0.4198

As part of a posthoc analysis to explore the relationship between arousal levels and performance, we defined two kinds of participants: high and low improvement groups in terms of their improvement score (IS) as denoted in Fig.42. Those participants whose IS was greater than the upper bound of the IQR (inter-quartile range), i.e., the top 25%, were defined to be in the high improvement group (N=14). Similarly, those participants whose IS was lesser than the lower bound of the IQR (the bottom 25%), were defined to be in the low improvement group (N=19). Table 14 shows the results of Mann-Whitney U tests to compare the physiological arousal metrics between these two groups. Among statistically significant differences, the mean SCRAmp of the low improvement group (0.25) was greater than that of the high improvement group (0.12) (p = 0.0298, U = 63), and the mean SC of the low improvement group (3.49) was greater than that of the high improvement group (1.59) (p = 0.0252, U = 55).

3.5. Discussion

This section discusses the results and is structured around each of the four research questions formulated in the related works section.

Is IVR training as effective as physical training in improving task performance?

Both IVR and physical training result in statistically significant improvements in contact time (CT) from pre-training to post-training phases. This shows that participants from both training conditions achieved fewer mistakes while performing the task. Participants in the IVR group showed improvements in task completion time which neared statistical significance. However, for participants who underwent physical training, the task completion time did not show a statistically significant change. Overall, the results suggest that training in fine motor skills results in quantifiable performance improvements for participants in both IVR and physical training. This is expected and as per the literature on IVR-based skill training (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021).

To compare the effectiveness of the two training modalities, three metrics to quantify improvement were defined: improvements in task completion time (TCT-I), improvements in contact time (CT-I), and an Improvement Score (IS) which combines the first two metrics. Statistical tests comparing these three metrics between IVR and physical conditions showed no statistically significant differences. Thus, the results indicate that IVR training is as effective as physical training for training in the buzz-wire task, thus supporting similar findings in other IVR skill training literature (Murcia-Lopez & Steed, 2018; Schwarz et al., 2020). One can argue that the novelty effect of IVR might have played a role in its effectiveness as it was observed that 31 participants in the IVR condition had tried VR only 1-5 times before the study, and 7 had never tried VR before. In a review, Merchant et al. (2014) found a link between the novelty effect of desktop VR-based high school education and learning outcomes and that the latter may even decrease as the number of VR sessions increases. Thus, novelty in VR use can play a role yet as the current study utilized a short familiarization task prior to the actual experimental task, this effect can only be a small attribute of the observed effectiveness

The current finding that IVR training is as good as physical training should also be considered in terms of the potential for further enhancement of this training modality. Literature suggests different methods to do this: the inclusion of haptic feedback (Frederiksen et al., 2020; Winther et al., 2020) and the inclusion of body representation and movements (other than the head and controllers) (Jensen & Konradsen, 2018). Inspirations for improving IVR training might also be taken from motor skill training literature which suggests techniques like decreasing the frequency of feedback as the skill level of the participant increases during training (Hebert & Coker, 2021), allowing participants to choose whether they want to receive feedback or not (Chiviacowsky & Wulf, 2005), or the IVR simulation adapting aspects of the training to the individual in real-time using physiological arousal levels and/or performance metrics (Zahabi & Abdul Razak, 2020).

Is there a significant difference in the enjoyment, presence, immersion, task load, and changes in self-efficacy reported by participants in IVR compared to physical training?

Participants in the IVR condition reported on average significantly more enjoyment levels than participants in the physical condition. This finding is consistent with IVR literature (Makransky et al., 2019). One parameter that is typically associated with frustration and lack of enjoyment during a VR experience is cybersickness. Herein, there were no incidents of cybersickness reported by the participants, probably due to the seated arrangement. Participants in the IVR condition reported on average more immersion (with statistical significance) than those in the physical condition. Similar trends exist for the presence measure, with participants in IVR training reporting more presence than those in physical training, where the difference was found to approach statistical significance. Though the IVR condition showed higher presence and immersion scores compared to the physical condition, it should be kept in mind that results from such metrics gain more importance when all subjects experience the same environment (Usoh et al., 2000). Nevertheless, the results are encouraging and as expected, as participants did not feel less immersed or present in the IVR environment as compared to the physical.

The NASA-TLX results show that in all parameters except temporal demand, IVR training induces roughly the same workload on participants as physical training. This was expected, as all kinds of visual noise and other confounding variables were tightly controlled across both conditions. However, VR, if not designed properly, may cause more cognitive load due to the possible complexity and novelty of the VR interactions involved. The one task load parameter where IVR training shows a statistically significant advantage over physical training is temporal demand. However, one cannot draw clear conclusions from this finding and further research is needed, for example, to compare the total training time across both conditions (which was not part of the research questions) along with the perceived temporal demand. This opens up interesting possibilities, due to the presence of a "time compression" effect in IVR as observed by Mullen and Davidenko (2021), where subjects experienced time to speed up while using VR compared to those in the control condition.

Participants in both physical and IVR training conditions reported an increase in self-efficacy, though a statistically significant increase was found only for the IVR group. Increases in self-efficacy levels have been found to correlate positively with learning outcomes (Makransky et al., 2019; Shu et al., 2019) and motor skill performance (Bandura, 1986). However, both VR and physical training in the current study did not show different levels of improvement in performance. It is possible that the novelty effects of VR caused participants in the VR condition to start initially with a lower self-efficacy in spite of the VR familiarization, but they ended up with self-efficacy levels similar to the physical condition by the end of the training. Further research is required to understand the links between self-efficacy and familiarity with the IVR medium. Additionally, these participants in the VR condition were observed to both have lower physiological arousal along with their increased self-efficacy. According to Bandura (1986)'s model of self-efficacy, there is a possible interaction between self-efficacy and arousal which merits further research in the context of IVR skill training.

Is there a significant difference between the physiological arousal levels of participants in IVR training compared to physical training?

Analysis of EDA and HRV metrics from the physiological arousal data revealed that IVR training caused less arousal than physical training, with a significant difference found for the SCRPeaks (EDA) metric and a near significant difference found for the HRV metrics Heart Rate and Inter-Beat Intervals. However, the frequency domain HRV measures, i.e., HFN, LF/HF ratio, and the time domain HRV measures SDNN and RMSSD showed no statistically significant difference.

Though there is no literature on the comparison of arousal between IVR and non-IVR conditions for skill training, some indicative literature from other domains exists. Tian et al. (2021) found more physiological arousal (EDA, EEG measures) in participants being emotionally stimulated through videos in the IVR condition as compared to those in the 2D condition. Egan et al. (2016) in a comparative quality of experience study found greater HR in the IVR condition compared to the non-IVR 2D condition while they found that EDA showed the opposite trend to our finding. We discuss possible causes for these seemingly contradictory trends towards the end of this section.

Is there a link between physiological arousal during training and improvements in performance after training?

A posthoc analysis was performed to compare the physiological data from participants with the highest improvement to those with the lowest improvement. This revealed greater arousal in two EDA measures (mean amplitude of skin conductance responses and mean skin conductance) for those participants who improved the least as compared to those who improved the most. This result is in alignment with findings from the literature; for example, in surgical simulation training (non-IVR), it has been found that lower performance is correlated with increased stress (higher arousal) levels (Prabhu et al., 2010; Quick et al., 2017). When correlation analysis was performed to compare the different arousal metrics with performance metrics for the whole study sample, we found statistically significant but weak correlations for improvement in task completion time (TCT-I) and among two EDA metrics: mean amplitude of skin conductance responses during training (SCRAmp) and mean skin conductance (SC). Further research should investigate the link between performance and arousal for participants across all levels of performance improvement.

For the last two research questions (links between arousal and training condition, arousal, and performance improvements), we found significant differences only in EDA metrics but not in HRV. This might be because EDA is purely a measure of sympathetic activity, as skin conductance levels are not counteracted by the parasympathetic nervous system. On the other hand, heart rate activity is controlled by both the sympathetic system (which causes heart activity to increase) and the parasympathetic system (which causes heart rate activity to decrease back to the baseline) (Cacioppo et al., 2007). Some literature finds EDA measures to be superior in terms of measuring changes in arousal (Dawson et al., 2017), even above HRV (Healey & Picard, 2005).

3.6. Limitations

Motor skill learning literature indicates the possibility that short-term performance might misrepresent learning (Magill & Anderson, 2016). Although a distractor task (see section 3.4) was used in the current study to compensate for the short-term nature of the retention test, it may be necessary to perform the retention tests after longer intervals to give a more precise understanding of the relationship between training conditions and retention. IVR skill training literature points to many comparative studies where retention tests after long intervals show better or the same retention in performance for the IVR condition as compared to non-immersive VR and physical conditions (Butt et al., 2018; Buttussi & Chittaro, 2018; Sakowitz et al., 2019). An illustrative example is in the burr-puzzle solving task by Carlson et al. (2015), where participants in a physical training condition initially outperformed those in IVR in terms of knowledge retention, but after two weeks, this effect was reversed. These examples suggest that such results may be expected in contexts similar to the current study, however further research is still required.

The study was also purposefully limited in terms of the 'training' provided. Here, participants were not given instructions during or after the training (knowledge of results) but participants get only automated feedback during the training when mistakes were made (knowledge of performance). Further research may build upon the design of the experiment and incorporate different training strategies or instructions. Also, the study is limited only to people who self-reported to be right-handed, to better control the setup and minimize variations, but future research might consider designing buzz-wire arrangements that are compatible with left-handed participants.

Regarding considerations on arousal metrics, comparisons using HRV metrics in the current study showed a lack of significant results. This could potentially be explained if it is assumed that the main cause for arousal in the current task was contact feedback (audio-visual-haptic). Since the time spent by a participant in contact with the wire (i.e., committing mistakes) will only be a proportion of the total duration of the training, any short-term increases in HRV metrics (which is accompanied by a rapid return to normal) may get averaged out by variations in HRV metrics during the rest of the training where they do not make any mistakes. Another potential confounder, which could cause variation in HRV, is the physical aspect of the activity where the participant has the freedom to choose any possible configurations of hand-arm-shoulder movement to complete the task with their right hand. Controlling this was beyond the scope of the current setup. Future studies may require a more fine-grained analysis of the relation between different stimuli (feedback during mistakes, difficulty in navigating certain parts of the wires) and physiological signals. Inspirations from the literature include Liebold et al. (2017) where a post-stimuli window of 10 seconds was used for heart rate metrics and

Boucsein (2012) which recommends a 1-5 seconds post-stimuli window to detect event-related skin conductance responses (ER-SCRs).

Regarding the choice of sensors used, the study is limited to only two measures of physiological signals (EDA and HRV). There is a multitude of physiological sensors which can be used to detect physiological arousal like Electroencephalogram (EEG), skin temperature, and eye-tracking. Additional sensors were not used as they might have made the experimental procedure more complex and affected the behaviour of the participants. However, additional sources of bio-signals merit further exploration in IVR training research as there are indications that some signals may make others redundant, for example, pupil dilation (from eye-tracking sensors) has been found to be correlated with both EDA and HRV (Wang et al., 2018). It is known that melatonin (which is correlated with the time of day), and temperature affect HRV and EDA metrics (Boucsein, 2012; Schachinger et al., 2008), but these factors were not controlled for in the experiment. On the other hand, these effects may have been reduced by the baseline correction applied to the various arousal metrics. Though arousal in this study is averaged across all the training sessions, the long recovery periods lasting several minutes for HRV signals to return to baseline levels (Moses et al., 2007), might potentially result in arousal from one level of training affecting the next. However, this issue may not affect EDA metrics, as a half recovery period from 2 to 10 seconds is found in the literature (Dawson et al., 2016), which is within the range of the 30 second rest interval between each level. The current study did not control for colour-blindness, and the self-reported normal vision of the participants was not medically certified, both of which might have caused differences in performance between the conditions.

A related factor affecting our study is the inherent difference between the haptic feedback available in the IVR and physical conditions. Though the vibration aspect is identical in both conditions, in the physical condition there is the added feel of the physical wire though the vibration masks this feeling to a certain degree. We propose further experimentation in IVR modality alone, with conditions being varied for various haptic feedback modalities like portable, grounded, and wearable as observed by Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al. (2021) in their analysis of the use of haptics in industrial skills training. The investigation of possible links between haptic feedback modality, physiological arousal, and improvements in performance holds promise for improving the state of the art in IVR-based skills training.

3.7. Implications for researchers

Taking as a point of departure the findings and lessons learned from this study one may consider:

- IVR and other training modalities must be designed to minimize distractions. This study tries to achieve this by using black panels covering the peripheral view of the participant and using headphones which, in addition to providing audio feedback, also minimizes external noise. In their review of motor skill learning literature, Wulf et al. (2010) found that performance is increased when there is an 'external focus' directed at the effect of the movement itself instead of an 'internal focus' directed at the trainee's body movements. Therefore, it is recommended that such complexities be minimized unless there are reliable methods of representing hands, arms, and other relevant parts of the body realistically. The coherence principle from the Cognitive Theory of Multimedia Learning further supports this by stating that removing stimuli irrelevant to the training context can improve learning outcomes (Parong & Mayer, 2021).
- VR hardware – The use of the Oculus Quest often requires minor calibrations related to the setting of tracking boundaries. This may be avoided by making sure the study environment is consistent between sessions or by using external trackers.
- Polar H10 – This cost-effective yet highly accurate and reliable ECG heart monitor is a useful tool for measuring arousal levels (Polar Electro Oy, Kempele, Finland). Researchers should, however, take into consideration the time taken for setting up the device and for the study setup to give privacy and instructions to participants for properly wearing the device.
- Shimmer GSR+ – This is a cost-effective and reliable device for measuring electrodermal activity (EDA) (Shimmer Research Ltd., Dublin, Ireland). The opportunity of measuring high-quality EDA signals from the fingers also restricts the training task from involving bimanual skills (use of both hands). Alternative but less accurate/convenient locations on the body can be considered if a training task demands the use of both hands (van Dooren & Janssen, 2012).
- Buzz-wire task – This task allows for one-hand use making it convenient for studies using EDA. The training task itself provides immediate feedback and allows for variations, for example, different types of audio, visual, or haptic feedback.

3.8. Conclusion

The study suggests that for the fine motor skill training presented, IVR training is as effective as physical training in improving task performance. Participants in the IVR condition reported an improvement in self-efficacy and significantly more enjoyment and immersion than physical training. Also, participants in the IVR condition on average displayed lower arousal than physical training. Though clear indications on the relationship between arousal and improvements in performance could not be found, EDA metrics hold potential for further investigation to answer this question by showing differences in arousal between high and low improvement groups. It is our understanding that such findings add to the IVR training field and can potentially pave the way to user-adaptive training systems (Zahabi & Abdul Razak, 2020).

Future work could incorporate subjective measures of arousal (like the Self-Assessment Manikin) into the immersive VR training as an additional layer to confirm findings from the physiological arousal signals. Additional measures like EEG could be employed to investigate the effect of the different types of stimuli on different brain regions, resultant cognitive load, and their relationship with arousal and performance (Hofmann et al., 2021; Tian et al., 2021). However, this should be implemented in a manner that does not break immersion/presence. It should also be noted that the current study does not explore the origins of the physiological arousal observed during the study but only its effects on performance improvement. It is reasonable to assume that the arousal observed may have been primarily caused by the direct feedback provided (visual, audio, and haptic) but other factors may also play a role. The study tries to control such extraneous factors by features in the study design like providing an initial baseline phase for the users to relax and also rest periods between training levels. The present study does not go into a fine-grained analysis of the relationship between arousal and stimuli like feedback from mistakes or challenging parts like bends in the wire, but rather looks at arousal across the whole training phase. There could be merit in understanding the short-term changes in arousal for various kinds of stimuli; for example, haptic feedback which is increasingly becoming a major focus point for IVR research as it affects task performance and presence (Kreimeier et al., 2019) and is crucial for many fine motor skill training tasks in VR like surgery (Rangarajan et al., 2020). This study also considers averaged performance metrics across the entire training session to answer the primary research questions, but future work might consider variations during the motor skill training, particularly in understanding different control strategies and stages of learning (Sternad, 2018). Future studies may also try to incorporate a crossover study methodology in order to control for difference between groups, by exposing the same group of participants to counterbalanced exposures to VR and physical training with appropriate time intervals in between to reduce cross-over effects similar to Yin et al. (2019).

Chapter 4 - Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task

This chapter details the investigation of the relationship between fine motor skill training in VR, haptic feedback, and physiological arousal. To do so, we present the design and development of a motor skill task (buzz-wire), along with a custom vibrotactile feedback attachment for the Geomagic Touch haptic device. A controlled experiment following a between-subjects design was conducted with 73 participants, studying the role of three feedback conditions -- visual/kinesthetic, visual/vibrotactile and visual only - on the learning and performance of the considered task. Results indicate that performance improved in all the three feedback conditions after the considered training session. However, participants reported no change in self-efficacy and in terms of presence and task load (NASA-TLX). All three feedback conditions also showed similar subjective arousal levels, but participants in the visual/kinesthetic feedback condition showed higher physiological arousal compared to those in the visual/vibrotactile one. Further analysis revealed that higher improvements in performance was linked to higher arousal levels. These results suggest the potential of haptic feedback to affect arousal levels and encourage further research into using this relationship to improve motor skill training in VR.

4.1 Introduction

Virtual Reality (VR) training is gaining popularity in medicine, rehabilitation, and industry, addressing various psychomotor, procedural, spatial, and decision-making skills (Abich et al., 2021; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021). VR training provides virtual environments where trainees receive consistent and replicable training, allowing for an objective assessment of skills. However, evidence for supporting the advantages of VR training over other methods is mixed. For example, in a Cochrane meta-analysis of VR-based endoscopy training literature, Khan et al. (2019) found that though VR training is better than no training, it is not better than conventional training, whereas Mekbib et al. (2020) in a meta-analysis of upper limb rehabilitation literature found that VR was better than conventional therapy. The effectiveness of VR for training may be enhanced by improving the sensory fidelity of VR, i.e., the “immersion” provided by the system through the use of head mounted displays (HMDs) or CAVE Automatic Virtual Environments (CAVEs) (Slater, 2018). A more immersive and interactive VR system may indeed lead to higher perceived presence (the subjective response to immersion) (Slater, 2018), which in turn positively affects the effectiveness of VR training (Makransky & Petersen, 2021). For the same reasons, haptic feedback in VR may increase presence and potentially training performance (Howard et al., 2019; Kreimeier et al., 2019).

However, employing haptics remains relatively unexplored in motor skill training literature in immersive VR (IVR) outside the surgical and rehabilitation domains. Therefore, further investigation, not only on the haptic feedback, but also on its modalities and variations can aid the discussion of how VR training might be enhanced (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021). Inspired by examples of motor skill training from immersive VR, (Christou et al., 2018; Radhakrishnan, Chinello, et al., 2022), this study focuses on a buzz-wire (or wire loop) task, where the aim is to move a metallic loop across a wire without touching it. This setup is used to investigate the following research questions:

- RQ 1: Can vibrotactile or kinesthetic feedback influence VR training performance in a buzz-wire motor skill task?
- RQ 2: Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in arousal levels during training? Is there a link between physiological arousal during training and improvements in performance afterward?
- RQ 3: Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in reported presence, task load, and self-efficacy?

4.2 Related Works

Haptics in immersive VR training

In a systematic review and meta-analysis of the use of different types of haptic feedback on VR and box trainers in laparoscopic surgical skills, (Overtoom et al., 2019) found that the addition of haptics provides only a small positive effect on task performance while providing a better learning curve at the beginning of training as compared to no haptics conditions. Similarly, Rangarajan et al. (2020) found that haptics enhanced surgical training further than training without haptics. They also found that the addition of haptic feedback reduced the learning curve for novice trainees. In fact, different modalities of haptic feedback have been documented in literature such as kinesthetic, vibrotactile and mid-air haptics. In this paper, we focus on kinesthetic and vibrotactile feedback modalities as they are suitable for providing relevant information regarding mistakes in the buzz-wire task. Kinesthetic feedback is closest to recreating real-world physical forces using grounded haptic devices, however, it faces the issue of relatively high cost of equipment and programming complexity to efficiently simulate haptic interaction. On the other hand, vibrotactile feedback is a form of cutaneous feedback that uses vibrations to convey virtual contact sensations, which is commonly bundled in commercial VR controllers like the Oculus Touch and the Vive controller.

VR training literature has compared vibrotactile and kinesthetic feedback against each other and other modalities. For example, (Islam & Lim, 2022) in their systematic review of vibrotactile feedback for motor skill training in VR found that vibrotactile feedback, used either alone or along with other feedback modalities was effective in most examples they analysed in VR literature. In the case of kinesthetic feedback, an indicative study is (Carlson et al., 2015) where IVR training with kinesthetic haptic feedback was compared to physical training of a burr puzzle assembly task. In the same work it was found that although physical training led to better immediate outcomes in terms of task time completion in tests, the VR group showed better performance on a delayed post-test. In a between-subjects comparison of different haptic modalities (vibrotactile and kinesthetic) for powered tool simulation, (Homer et al., 2019) found that vibrotactile combined with audio feedback led to the greatest increase in performance, whereas the addition of kinesthetic feedback did not improve performance. Kreimeier et al. (2019) performed a within-subjects study comparing the effect of vibrotactile feedback (plus visual), kinesthetic feedback (plus visual), and visual only (no haptics) on a throwing task in IVR, finding that kinesthetic feedback led to better task performance compared to the other modalities. These examples from literature illustrate the need for further research into investigating the effectiveness of different haptic feedback modalities.

Haptic feedback and physiological arousal

The term “arousal" refers to the increase in alertness and attention in response to external or mental stimuli. Subjective methods to measure arousal include questionnaires like the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) and the Affective Slider (AS) (Betella & Verschure, 2016). Physiological measures of arousal are obtained by measuring signals from the autonomic nervous system (ANS) including EDA (electrodermal activity) caused by sweating in response to arousal, HRV (Heart Rate Variability), respiration volume/rate, pupil diameter variation, and brain activity (Cacioppo et al., 2007). EDA is a common choice for measuring ANS activation due to its neuroanatomical simplicity (Dawson et al., 2016). HRV is also popular as it enables the differentiation of various psychological and physiological states (Dawson et al., 2016). The relationship between arousal and task performance has been investigated in the literature. Though it has been hypothesized to be linked to performance in an inverted U-shaped curve according to Yerkes-Dodson law (Yerkes & Dodson, 1908), this has not been conclusively established by literature due to other factors like task complexity and personality factors (Bargh & Cohen, 1978; Storbeck & Clore, 2008). Increases in arousal for example can affect memory and cause retrieval of task-irrelevant information, which may affect training outcomes (Plass & Kalyuga, 2019). Though as discussed, both haptic feedback and arousal may affect task performance, the relationships between these three factors are not well established. There are few explorations linking arousal to haptic feedback in literature, for example, Gatti et al. (2013) used a Geomagic Touch (3D Systems, US) to render viscous forces onto participants' hands while they viewed emotional pictures where they found an effect of haptics on subjective arousal (SAM) but not on physiological arousal (EDA, HRV, respiratory rate, and temperature). (Sampath et al., 2015) similarly linked an emotional pictures dataset with vibrotactile haptics and subjective arousal (SAM), finding that high-intensity haptic feedback on the fingers contributes to increases in subjective arousal. The literature on haptic feedback and physiological arousal is not well established in the VR literature, except for Krogmeier et al. (2019) for a non-motor skill training context and two pilot studies in the motor skill training domain (Koumaditis et al., 2018) (Radhakrishnan, Koumaditis, et al., 2022).

In addition to measures of performance and arousal, subjective measures of presence, and task load among others add additional insight into designing more effective haptics-enabled VR training. There are examples from the literature of investigations into the relationship between presence and haptic feedback, where for instance Gibbs et al. (2022) and Cooper et al. (2018) found that multimodal (i.e., combinations of haptics with audio or visuals) feedback led to better presence as compared to providing feedback in any one modality alone. However, it is not clear which feedback modality, between kinesthetic and vibrotactile can cause the greatest increase in presence. Additionally, research has pointed to the links between task load and haptic modality. For example, kinesthetic feedback has been linked to lower task load in surgical VR training (Zhou et al., 2007) and VR motor skills therapy scenarios (Ramírez-Fernández et al., 2015). Weber et al. (2013) in a desktop VR peg-in-hole experiment found that kinesthetic feedback led to a lesser overall task load compared to vibrotactile and visual (no haptic feedback) conditions. It remains an open question if this pattern holds for the buzz-wire task in immersive VR which requires finer motor skill control. Self-efficacy is another subjective measure that has been linked to motor skill performance (Bandura, 1986). Though studies have found VR training to be linked to increases in self-efficacy (Radhakrishnan, Chinello, et al., 2022; Shu et al., 2019), its link to training outcomes is unclear.

4.3 Methods

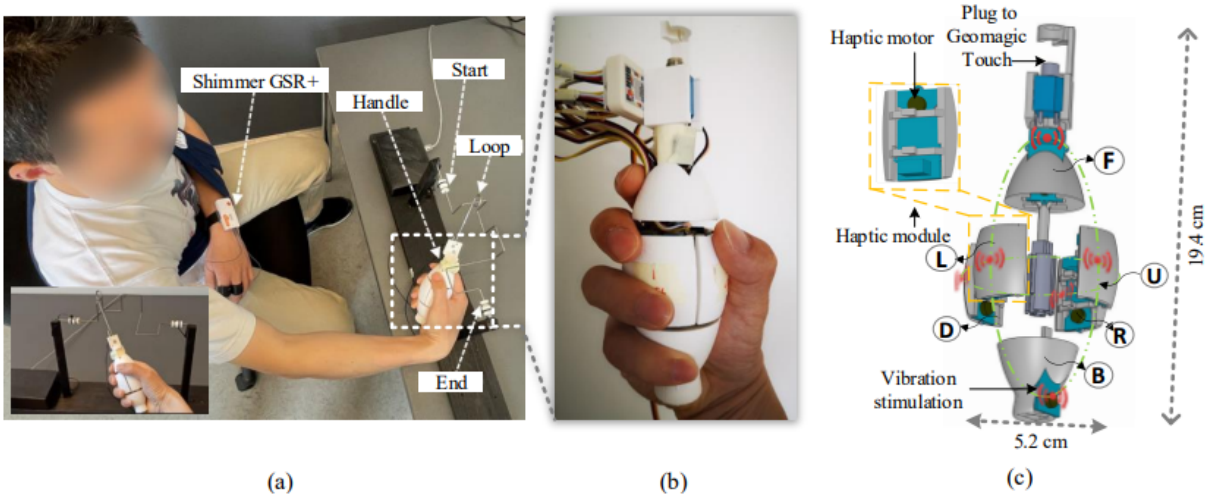


Figure 43 Experimental setup (physical tasks). (a) Physical experimental environment, where the participant moves the loop across the real wire. The participant wears an Electrodermal Activity (EDA) sensor (Shimmer GSR+). The Heart Rate Variability sensor (Polar H10) is worn around the chest, in contact with the skin (not visible in the picture). (b) The custom handle held by a participant. It houses six vibrotactile motors able to provide distributed vibrations when a mistake happens, i.e., the wire touches the loop. It was attached to a real metallic loop during the physical tasks (as in (a)) and to a Geomagic kinesthetic interface during the VR tasks (as in Fig. 45). (c) CAD representation of the custom handle, highlighting the positioning of the vibrotactile actuators, L: left, U: up, R: right, D: down, F: front, B: back.

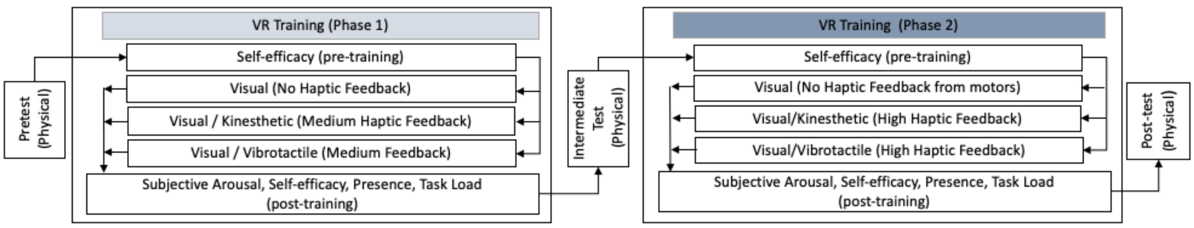


Figure 44 Flow of the experiment. First, participants perform the buzz-wire task in the physical environment ("pretest"), where they hold the handle attached to a metal loop and interact with a real buzz-wire. Then, participants are given training in an immersive VR environment ("VR Training (Phase 1)"). This time, participants hold the handle attached to a Geomagic Touch haptic interface and interact with a virtual buzz-wire, receiving either visual feedback only, visual and kinesthetic feedback, or visual and vibrotactile feedback. After this VR training, participants are asked again to perform the buzz-wire task in the physical environment ("intermediate test"). The experimental protocol is then repeated, with the participants again receiving training in an immersive VR environment ("VR Training (Phase 2)"), and then carrying out a final buzz-wire physical task ("post-test"). The two phases of VR training ("Phase 1" and "Phase 2") provide the participants with different intensities of feedback. This experimental organization enables us to study the effect of VR training in the learning of the considered fine motor skill task. To do so, before the training, we asked participants to estimate their perceived efficacy in carrying out the task ("pre-training"), and, after the training, we asked participants to estimate their perceived task load, self-efficacy, subjective arousal, and presence ("post-training"). Other metrics, such as task completion time, contact time with the wire, and physiological arousal, are meausred throughout the whole experimental session.

We designed an experiment to address the research questions presented at the end of Section 4.1. As a representative example of fine motor skill task, we considered a buzz-wire (or wire loop) task, where the aim is to move a metallic loop across a wire without touching it (see Fig. 43a). The flow of the experiment is shown in Fig. 44, from left to right. First, all participants are asked to perform the buzz-wire task in a physical environment, called "pretest" in Fig. 44. During this task, participants hold a handle attached to a metal loop (see Fig. 43), which they move from one end of a physical wire to the other, as fast as possible and with the least number of mistakes. Then, participants are given training in an immersive VR environment to improve their task performance, called "VR Training (Phase 1)". The VR training environment consists of multiple virtual buzz-wires, similar to the one in the pretest environment, but featuring different levels of difficulty (see Fig. 45). This time, participants hold the handle attached to a Geomagic grounded haptic interface. Similarly, as before, they had to move the loop across the considered virtual buzz-wire as fast as possible and with the least number of mistakes. During this VR training, participants receive different types of feedback about the contacts of the loop with the wire, according to the group they have been assigned to: one group of participants receives visual feedback only, one receives visual and kinesthetic feedback, and one receives visual and vibrotactile feedback. After this VR training, all participants are asked again to perform the buzz-wire task in the physical environment, called "intermediate test". This post-training task enables us to compare how the VR training affected user's performance according to the feedback provided. Finally, this experimental protocol is repeated, with the participants once again receiving training in an immersive VR environment, called "VR Training (Phase 2)", and subsequently analyzing the change in their performance in a final buzz-wire physical task, called "post-test" (see the right-hand part of Fig. 44). The two phases of VR training ("Phase 1" and "Phase 2") provide the participants with different intensities of feedback, enabling us to also analyze whether the intensity of the feedback affect the user's performance.

Performance of the task is measured using both objective metrics (completion time, mistakes, physiological arousal) and subjective measures (presence, task load, self-efficacy).

The following sections detail each of the aforementioned parts of our experiment.

Physical experimental environment

The physical environment is shown in Fig. 43. It is an adaptation of the physical test of motor skills used by (Radhakrishnan, Chinello, et al., 2022). The metal wire is long 52~cm and rests on two 20~cm-high pillars. The start and end positions are denoted by grooves on two white plastic cylinders on either side of the wire, designed so that the loop can rest on them. An Arduino UNO is used to detect the mistake signals, i.e., when the loop touches the wire, and transmit them to an external computer through a serial connection. Similar circuits are used to detect when the loop is lifted off the groove at the beginning of the task as well as when the loop reaches the groove at the end of the wire. A video showing the physical setup can be found at https://youtu.be/qZ6fBP-poAs?t=106. Participants were asked to carry out the task in this physical environment three times throughout the overall experience ("pretest", "intermediate test", and "post-test"), as described before and summarized in Fig. 44.

Custom vibrotactile handle

Figure 43 shows the custom handle designed to provide vibrotactile sensations about the (undesired) contacts between the loop and the wire, according to the feedback condition at hand. It is shaped as an ellipsoid and houses six vibrotactile modules, inspired from (Aggravi et al., 2018; Cabaret et al., 2022). The vibrotactile modules are positioned around the handle: four are placed symmetrically around the plane perpendicular to the main axis, and two are placed at the ends of the main axis (see Fig. 43). Small gaps around where the haptic modules are positioned weaken the transmission of vibrations, making it easier to recognize the source of the vibration.

We carried out a perceptual experiment, enrolling 12 participants, to evaluate the capability of the handle to provide spatialized vibrotactile sensations. We activated one random motor at a time and asked the participants to indicate which one was vibrating (30 trials in total, 5 per motor). Participants were able to correctly recognize the activated motor 68.0\% of the times (see Fig. 46). On the other hand, 13.6% of the times participants confused adjacent motors, while 18.4% they confused motors located further away.

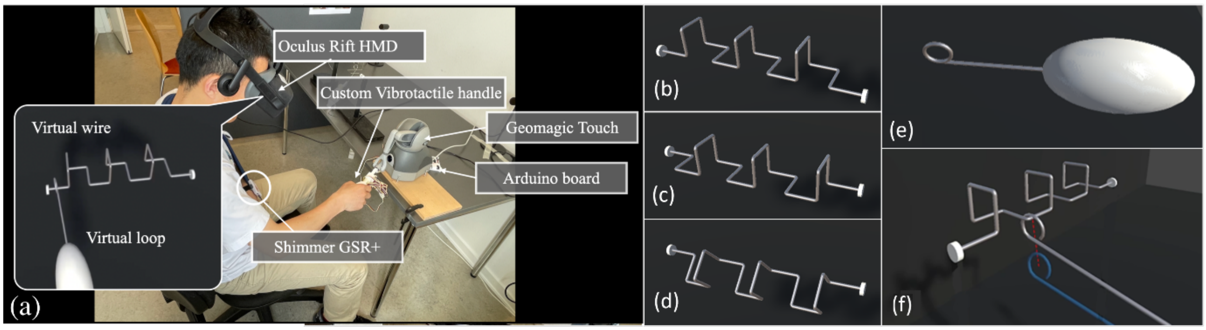


Figure 45 Experimental setup (Virtual Reality training). (a) The Virtual Reality (VR) training environment, where participant holds the custom handle attached to the Geomagic Touch haptic device. (b) Level 1, (c) level 2, and (d) level 3 of the buzz-wire task, presented during the two phases of the VR training. (e) Detail of the the loop and handle avatars in VR, which have the same dimensions of their physical counterparts. (f) Visual feedback provided to the users when the loop contacts the wire: a semi-transparent blue loop indicates the true position of the loop as commanded by the user, while the standard opaque grey loop indicates the proxy/ideal position of the loop inside the wire. Visual feedback is provided in all the feedback conditions.

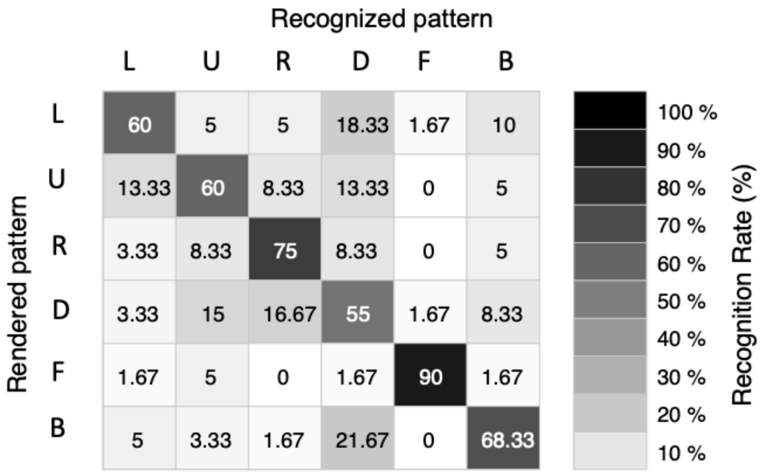


Figure 46 Custom vibrotactile handle: perceptual experiment. Confusion matrix showing the recognition rates when activating each of the six motors (see also Fig. 43).

VR training environment

The VR training environment is shown in Fig. 45. It was composed of three virtual buzz-wires levels, shown in Figs. 45b, 45c, and 45d. Level 1 was 57~cm long (end to end) with a horizontal span of 21~cm and eighteen 90º bends (see Fig. 45b); level 2 was designed to be the mirrored version of level 1 so that the beginning and the end were inverted (see Fig. 45c); and level 3 was designed to be the same as level 1 but tilted 45º around the main axis of the wire (Fig. 45d). Participants were asked to wear an Oculus Rift Head Mounted Display (HMD) and hold the custom handle attached to a Geomagic Touch interface, as shown in Fig. 45a. The latter is used to track the movement of the user to animate the virtual loop in all feedback conditions (see Figs.45e and 45f). It is also used to provide kinesthetic feedback in the dedicated feedback condition.

Feedback conditions during VR training

Participants undergo two training phases in VR to become better at the buzz-wire task. They are randomly assigned to one of the three VR training conditions: visual feedback only, visual, and kinesthetic feedback, and visual and vibrotactile feedback about the contacts between the loop and the wire, that users are asked to minimize. Each subject carries out the VR training in only *one* feedback condition.

Across the two VR training phases ("Phase 1" and "Phase 2", see Fig. 44) the intensity of the haptic feedback, kinesthetic and vibrotactile, changes. Haptic feedback provided during Phase 2 is 50\% stronger than that provided during Phase~1, for kinesthetic and vibrotactile feedback, respectively. On the other hand, visual feedback does not change across the two VR training phases.

We carried out a short preliminary experiment to ensure that this difference in haptic intensity was noticeable. The change was indicated as "clearly noticeable" by the 12 participants of the preliminary test, who were asked to carry out the VR buzz-wire task four times, one per feedback intensity (medium, high) and type of haptic feedback (kinesthetic, vibrotactile), and describe their experience. The indications and results of this preliminary study have been used to set the parameters of the provided feedback as well, as detailed in the following Sections.

Below we detail the three types of feedback conditions.

(1) *Visual feedback only*
In this feedback condition, whenever the loop touches the wire, the virtual representation of the loop doubles, as shown in Fig. 45f and at https://youtu.be/qZ6fBP-poAs?t=65. A semi-transparent blue loop indicates the true position of the loop as commanded by the user, while the standard opaque grey loop indicates the proxy/ideal position of the loop, still inside the wire. It is worth noticing that here both the motors in the custom vibrotactile handle and the kinesthetic force feedback provided by the Geomagic Touch actuators are not active, while the user still holds the same handle attached to the kinesthetic interface to perform the task. During the contact situation, the opacity of the semi-transparent blue loop increases as it moves away from the proxy position of the loop inside the wire. Finally, a dotted line indicates the direction where the user should move to rejoin the wire and continue the task; the

color of the line changes from black to red as the user moves the loop away from its proxy position inside the wire.

(2) *Visual and kinesthetic haptic feedback*
In this case, whenever the loop touches the wire, participants receive the visual feedback as well as kinesthetic feedback forces provided by the Geomagic Touch interface. Specifically, the loop-wire haptic interaction was rendered using a simple elastic model with stiffness 56.7 N/m and 85 N/m for the two phases of the VR training, respectively. The linear damping was kept constant in both phases at 8 Ns/m. These values were chosen following the indications of the preliminary study, so as to resemble as much as possible the interaction with the physical buzz-wire.

(3) *Visual and vibrotactile haptic feedback*
In this case, whenever the loop touches the wire, participants receive the visual feedback as well as vibrotactile feedback stimuli provided by the custom handle. Specifically, vibrations were provided along the direction where the contact between the virtual loop and wire happened, e.g., if the loop touched the wire in its upper sector (as in Fig. 45f), the vibration was provided by the motor U (up) (see Fig. 43c); if the loop touched the wire in its lower sector, motor D (down) would be activated. The vibration amplitude was fixed and set to 0.33g for phase 1 and 0.49g for phase 2 of VR training. These values were chosen again following the indications of the preliminary study, to resemble as much as possible the interaction with the physical buzz-wire.

Metrics

Objective performance metrics

In their review of haptic feedback for motor skill training, (Basalp et al., 2021) listed three types of metrics that can be used to objectively measure performance, i.e., spatial, temporal, and spatiotemporal metrics. We considered three metrics:

a) Task completion time (TCT), which is the time taken to move the loop from start to end;
b) Loop-wire contact time (CT), which is the time spent by the wire in contact with the wire.
c) Improvement score (IS), which is calculated by rank ordering improvements among all participants into ten equal quantiles for TCT and CT separately. A rank order of 1 denotes the least improvement in performance while a rank order of 10 denotes the most improvement. Subsequently, IS for a participant is defined as the sum of the ranks

for TCT and CT, e.g., a participant who has improved the most in both TCT (rank order = 10) and CT (rank order = 10) would get an IS of 20.

Subjective measures

We considered subjective metrics related to perceived task load, self-efficacy, subjective arousal and presence. To measure task load, we used NASA-TLX(Hart & Staveland, 1988), across the six standard dimensions - Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. A combined task load score was then taken by averaging all six dimensions. To measure self-efficacy, the participants are asked the question *"How confident are you that you can perform a similar task effectively (go from start to finish as fast as you can with minimal mistakes) on a scale from 1 to 7?"*, both before and after the training (see also Fig. 44). To measure subjective arousal, we combined the arousal sub-scale of Self-Assessment Manikin~(SAM) (Bradley & Lang, 1994) with the Affective Slider (Betella & Verschure, 2016), as seen in (Granato et al., 2018). To measure presence, we used the five dimensions in the physical subscale of the Multimodal Presence Scale (Makransky et al., 2017).

Physiological arousal metrics

To investigate the participants' arousal levels, heart rate variability, and electrodermal activity, we used the Polar H10 (Polar Electro Oy, Finland) and Shimmer GSR+ (Shimmer Research Ltd., Ireland) sensors. Baseline values were subtracted where applicable from the physiological arousal metrics to control for individual physiological differences (Braithwaite et al., 2013). All physiological sensor data were streamed and stored using the iMotions platform (iMotions A/S, Denmark). See Section 4.2 for details about how these metrics are used in the literature.

Table 15 Relationship between increases in physiological arousal and EDA and ECG metrics.

Effect of Arousal (↑: increase, ↓: decrease)	
EDA	Skin Conductance (SC) ↑
	Skin Conductance Response Amplitude (SCRAmp) ↑
	Skin Conductance Response Peaks Rate (SCRPeaks) ↑
ECG	Heart Rate (HR) ↑
	Inter-Beat Interval (IBI) ↓
	Root Mean Square of Successive Differences (RMSSD) ↓
	Standard Deviation of NN Intervals (SDNN) ↓
	Normalized High-Frequency Component (HFN) ↓
	LF/HF (Low Frequency/High Frequency) Ratio ↑

a) *Heart Rate Variability*
The Polar H10 is an Electrocardiogram (ECG) heart rate monitor often used in VR studies (Radhakrishnan, Chinello, et al., 2022). The sensor is worn around the chest with direct contact with the skin. Data in the form of heart rate and Inter-beat Intervals (R-R intervals) are transmitted via Bluetooth at a rate between 1 and 2 Hz to the PC and then recorded by the iMotions application. Based on the raw R-R interval data, different measures of heart rate variability, including time and frequency domain metrics, were calculated using the hrv-analysis Python library (https://github.com/Aura-healthcare/hrv-analysis). Table 15 shows the link between increases in arousal and its effect on ECG based metrics. Increases in arousal are indicated by increases in Heart Rate and Low Frequency/High Frequency (LF/HF) Ratio (Orsila et al., 2008). On the other hand, decreases in HRV measures like IBI (Inter-Beat Interval), SDNN (Standard Deviation of NN Intervals), RMSSD (Root Mean Square of Successive Difference), and the frequency domain measure HFN (Normalized High-Frequency Component), indicate an increase in arousal (Narciso et al., 2020).

b) *Electrodermal activity (EDA)*
The Shimmer GSR+ (Galvanic Skin Resistance) measures EDA by passing a small current through electrodes placed on the index and middle fingers in the left hand (Radhakrishnan, Chinello, et al., 2022), to allow the participants to use their right hand for carrying out the buzz-wire task. The sampling rate was 128 Hz. EDA measures used in this study include SC (Skin Conductance), which increases in response to an increase in arousal (Collet et al., 2005). Table 15 also shows the link between increases in arousal and its effect on EDA metrics. An increase in arousal also leads to a higher rate of Skin Conductance Response Peaks per minute (SCRPeaks) which are peaks in the SC amplitude lasting between 1-5 seconds after onset (Krogmeier et al., 2019). Similarly, the mean peak amplitude of all SCR peaks (SCRAmp) is also a positive measure of arousal (Krogmeier et al., 2019). All EDA signals were processed using the Neurokit2 Python library (Makowski et al., 2021).

Experimental procedure and participants

Participants signed up for the study using the University online participant recruitment system and were randomly assigned to one of the three feedback conditions. They had to satisfy these criteria: (a) right-handed (to minimize variation in the setup), (b) normal vision or corrected-to-normal vision with contact lenses (no glasses), and c) no mental illnesses or sensitivity to nausea. Participants signed up for 35-minutes time slots of their choosing and were paid the equivalent of 12 Euros. Approval for this experiment was obtained from the Cognition and Behavior Lab's Human Subjects Committee (approval code: 339), Aarhus University.

After reading and signing the consent forms, participants were familiarized with the experimental task by an experimenter, who demonstrated the task. Thereafter, participants were given privacy to place the Polar H10 around their chest. Afterwards, the Shimmer GSR electrodes were placed on the index and middle fingers of the participant's left hand. Finally, participants were assisted in wearing a sling around the neck so that they could rest their left hand with the fingers relaxed, as seen in Fig. 43a. Participants were asked to keep their left hand still, so as to minimize the noise in the recorded signals. The signal quality for both sensors was verified before the experiment started. Baseline data from these biosensors were measured with the participants seated quietly, with their eyes closed, breathing normally, and without wearing the HMD.

The overall flow of the experiment is summarized in Fig. 44.

4.4 Results

Participants and data analysis

73 participants enrolled in the study, randomly divided between visual/kinesthetic (n=26), visual/vibrotactile (n=22), and visual only (n=25) feedback conditions. 37 participants identified themselves as female, 35 as male, and 1 as other. 36 participants indicated their age group in the 25-34 range, 33 in the range 18-24, and 4 in the range 35--44. More than half of the participants in the study (49) indicated that they had tried VR using a head-mounted display more than once, while 24 reported never having used it before the experiment. An analysis of the quality of the HRV and EDA data using iMotions led to dicarding of HRV and EDA data from 3 and 20 participants, respectively, due to signal quality issues. Unfortunately, this is a common problem for these measurements, as they are very sensitive to movements (EDA) (Boucsein, 2012) and the fastening/positioning of the sensor (EDA, HRV), which can change inadvertently during the experiment. The other metrics were not affected. Shapiro-Wilk tests for normality were applied to all the variables, and if a variable was found to violate

assumptions of normality, non-parametric statistical alternatives were used. A significance level of 0.05 was selected while interpreting the results of the statistical tests.

Objective performance

We analyzed the task completion time~(TCT), contact time~(CT), and improvement scores~(IS), as described in Section 4.3. Improvements in performance metrics due to training are evaluated as changes:

- from the pretest to the post-test physical tasks (referred to as "overall training");
- from the pretest to the intermediate test physical tasks (referred to as "Phase 1 training");
- from the intermediate test to the post-test physical tasks (referred to as "Phase 2 training").

Table 16 Summary of Two-way Mixed ANOVA Results for Task Completion Time and Contact time by Feedback Condition (between-subjects) and Test Iteration (within-subjects)

Dependent variable	Source of variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-value	p-value
Task Completion Time	Feedback condition	2883.949	2	1441.974	1.930	.153
	Test iteration	1060.274	1.473	719.692	7.582	.003*
	Feedback condition * Test iteration	181.935	2.946	61.747	.650	.582
Contact Time	Feedback condition	32.951	2	16.476	.992	.376
	Test iteration	95.662	2	47.831	18.113	<.001*
	Feedback condition * Test iteration	7.940	4	1.985	.752	.559

* denotes significant difference at α=0.05

Task completion time (TCT)

Fig. 47a shows the task completion time across the three physical buzz-wire tests. A two-way mixed ANOVA was conducted to analyse the differences between "feedback conditions" (between-subjects factor: visual, visual/kinesthetic, visual/vibrotactile) and "test iterations" (within-subjects factor: pretest, intermediate, and post-test physical tests) on this metrics. Table 16 summarizes the results of the following analysis. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2)=30.089$, $p<.001$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon=0.737$). There was no statistically significant interaction between the feedback condition and test iteration on TCT, $F(2.946, 101.653)=.650$, $p=.582$, partial $\eta^2=.019$. The main effect of time showed a statistically significant decrease in TCT over the three test iterations, $F(1.473, 101.653)=7.582$, $p=.003$, partial $\eta^2=.099$, with TCT decreasing from 31.53 ± 21.57s (mean ± standard deviation) in the pretest, to 27.98 ± 15.64s in the intermediate test and to 26.2 ± 13.97s in the post-test.

Post hoc analysis with a Bonferroni adjustment revealed that TCT decreased significantly from pretest to intermediate test (3.63s, p=.024), and from pretest to post-test 5.34s, p=.01) but not from intermediate test to post-test (1.7s, p=.304). The main effect of group showed that there was no statistically significant difference in task completion time between feedback conditions F(2, 69)=1.93, p=.153, partial η^2=.053.

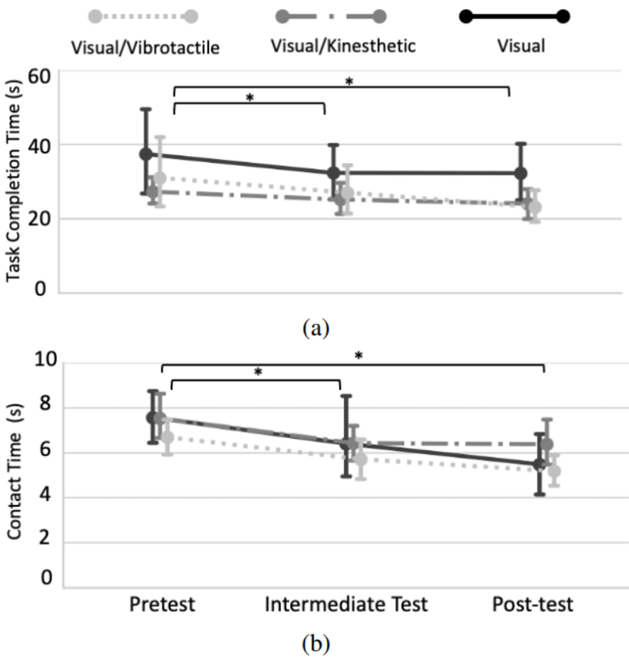


*Figure 47 Mean performance metrics with error bars (CI : 95%) across the three physical buzz-wire tests for Visual/Vibrotactile, Visual/Kinesthetic, and Visual feedback conditions for (a) task completion time (TCT) and (b) contact time (CT). * denotes significant difference at α=0.05.*

Contact time (CT)

Fig. 47b shows the contact time across the three physical buzz-wire tests. As before, a two-way mixed ANOVA was conducted to analyse the differences between feedback conditions and test iteration for this metrics (see Table 16 for a summary of the results). Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, χ^2 (2) = 4.754, p=.093. There was no statistically significant interaction between the feedback condition and test on CT, F(4, 138)=0.75, p=.559, partial η^2=.021. The main effect of time showed a statistically significant difference in CT over the test iterations, F(2,138)=18.11, p<.001, partial η^2=.208, with CT decreasing from 7.25 ± 2.49s (mean ± standard deviation) during the pretest, to 6.18 ± 2.99s in the intermediate test, and to 5.68 ± 2.58s in the post-test. Post-hoc analysis with a Bonferroni adjustment revealed that CT decreased significantly from pretest to intermediate test (1.07s, p=.002), and from pretest to post-test (1.61s, $p<.001$) but not from intermediate test to post-test (0.53s, $p=.153$). The main effect of feedback condition showed that there was no statistically significant difference in contact time between the feedback conditions, $F(2,69)=0.992$, $p=.376$, partial $\eta^2=.028$.

Improvement Score (IS)

A one-way ANOVA was performed to compare the improvement score between the three feedback conditions, which revealed no statistically significant difference ($F(2,69)=0.047$, $p=0.95$).

Self-Efficacy, Task load, Presence

A two-way mixed ANOVA was conducted to analyse the effect of feedback condition (between-subjects factor: visual, visual/kinesthetic, visual/vibrotactile)) and test iteration (within-subjects factor: pre-phase 1 training, post-phase 1 training, pre-phase 2 training, post-phase 2 training) on self-efficacy (SE). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 $(5)=32.45$, $p<.001$. Therefore, a Greenhouse-Geisser correction was applied ($\varepsilon=0.788$). There was no statistically significant interaction between the feedback condition and time on self-efficacy, $F(4.729,165.52)=1.143$, $p=.339$, partial $\eta^2=.032$. The main effect of time did not show a statistically significant difference in mean SE at the different time points, $F(2.365,165.52)=1.549$, $p<.0005$, partial $\eta^2=.022$. The main effect of feedback condition showed that there was no statistically significant difference in mean SE between feedback conditions $F(2,70)=1.135$, $p=.327$, partial $\eta^2=.031$.

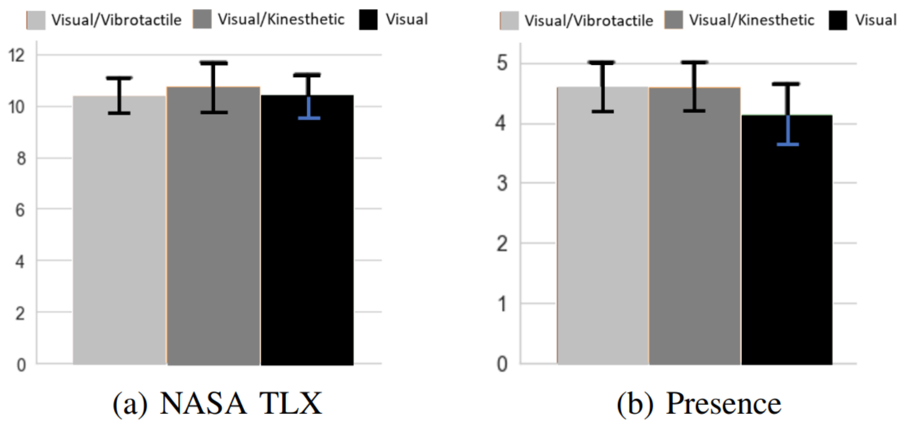


Figure 48 (a) NASA TLX and (b) Presence scores across Visual/Vibrotactile, Visual/Kinesthetic, and Visual conditions. Mean and 95% confidence interval are reported.

Fig. 48a shows the overall NASA-TLX score obtained by calculating the average of the six NASA-TLX task load dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration) reported by the participants. One-way ANOVAs/Kruskal-Wallis tests were performed to compare the effect of feedback condition on the individual NASA-TLX dimensions revealing no significant differences between the feedback conditions. However, posthoc pairwise comparisons (Mann-Whitney U tests) revealed a near statistically significant difference in temporal demand between visual/vibrotactile condition where participants reported on average a score of 10.02, and visual/kinesthetic condition where participants reported 11.94 (p=0.07, U=371.5).

Fig. 48b shows the overall presence score reported by the participants, calculated by taking the average of the responses to the five questions in the presence questionnaire. A Kruskal-Wallis test revealed no significant difference in the presence score between the feedback conditions (F(2,70)=1.5, p=0.22).

*Table 17 Physiological Arousal Metrics across High (n=15) and Low Improvement groups (n=16). * denotes significant difference at α=0.05*

Arousal Metric		High Improvement(Mean±SD)	Low Improvement (Mean±SD)	p-value
HRV	HR	4.82±5.59	-1.16±3.28	0.002*
	IBI	-42.27±55.81	4.7±34.69	0.02*
	RMSSD	-58.41±101.15	17.68±35.54	0.001*
	SDNN	-47.91±63.07	4.08±32.14	0.003*
	LF/HF Ratio	-1.47±3.97	-1.44±4.97	0.98
	HFN	1.33±17.83	7.41±14.89	0.35
EDA	SC	2.35±3.18	1.84±1.81	1.0
	SCRAmp	0.14±0.22	0.14±0.16	0.99
	SCRPeaks	10.37±4.76	8.84±1.84	0.36

Physiological arousal

Arousal and feedback condition

One-way ANOVAs performed on SAM (Self-Assessment Manikin) reported by participants did not reveal an effect of feedback condition on subjective arousal during training (H(2)=0.47, p=0.79). One-way ANOVAs performed on each of the EDA and HRV metrics also did not show any significant effect of feedback condition on those metrics. However, posthoc pairwise comparisons (Mann-Whitney U tests) showed that participants in the visual/kinesthetic condition had a higher baseline corrected LF/HF ratio (-0.15) compared to participants in the visual/vibrotactile condition (-2.59) (p=0.028, U=349). To measure differences in arousal immediately after the participants commit a mistake during VR training, the physiological arousal levels were averaged for 10-seconds windows starting at the moment of contact between the loop at the wire. One-way ANOVAs/Kruskal-Wallis tests performed to compare the effect of feedback condition on all of these "immediate" arousal metrics did not show any significant difference between the three feedback conditions.

Arousal and task performance

The link between task performance and arousal was assessed using co-relation tests between improvement score~(IS) and arousal, as measured by both subjective (SAM) and physiological (EDA, HRV) measures. Statistically significant (but weak) negative correlations were observed between Improvement Score (IS) and three HRV metrics: RMSSD (ρ=-0.35, p=0.003), SDNN (ρ=-0.32, p=0.008) and HR (ρ=0.245, p=0.043). As decreases in RMSSD and SDNN, and increases in HR, are correlated with increases in arousal (see Table 15), the results of the correlations tests show that improvements in performance are correlated with increases in arousal. To further investigate this point, we grouped participants into two groups: high-improvement (n=15) and low-improvement (n=15), respectively. Following the technique described by Radhakrishnan, 2022 #328}, the high-improvement group was defined as participants who had an improvement score (IS) above the upper range of the inter-quartile range (i.e., above the 75\% mark) and the low-improvement group had participants from the lower quartile range (i.e., below the 25\% mark). As shown in Table 17, Mann-Whitney U tests were performed for HRV and EDA metrics between these two groups. It can be observed that participants in the high-performance group had a higher baseline adjusted heart rate (HR) of 4.82, as compared to -1.16 of the low-performance group (p=0.002, U=178). In terms of inter-beat interval (IBI), the high-performance group had a lower baseline-adjusted IBI (-42.27) compared to the low-performance group with an IBI of 4.7 (p=0.02, U=51). Participants in the high-performance group showed a lower baseline adjusted RMSSD (Root Squared Mean of Successive Differences) of -58.41 compared to the RMSSD of 17.68 for the low-performance group (p=0.001, U=30). Similarly, participants in the high-performance group also showed a lower baseline-adjusted SDNN (Standard Deviation of NN intervals) of -47.91, compared to an SDNN of 4.08 for the low-performance group (p=0.003, U=37). This result supports the patterns observed in the correlation tests, i.e., improvements in performance are correlated with increases in arousal.

To further investigate the trends in the relationship between arousal and performance, participants were also divided into high-arousal and low-arousal groups regarding, separately, the HRV and EDA metrics. This split followed the same principle we used for splitting the participants into high and low-performance groups, i.e., according to their position in the inter-quartile range. Subsequently, the improvement scores (IS) were compared for each of these pairs of high and low-arousal groups using Mann-Whitney U tests. In terms of high and low-arousal groups split according to RMSSD, participants in the high-arousal group had an IS of 9.06, which was lower than the IS of 12.06 observed in the low-arousal group ($p=0.01$, $U=66$). Similarly, for groups split according to SDNN, participants in the high-arousal group were observed to have an IS of 9.06, which was lower than the IS of 12.31 observed in the low-arousal group ($p=0.017$, $U=70$). Among groups split according to HR, participants in the high-arousal group had an HR of 12.39 as compared to the HR of 9.78 observed in the low-arousal group, and this difference approached the threshold of statistical significance ($p=0.09$, $U=221$). This further confirms the trends observed earlier of performance improvement as arousal increases.

4.5 Discussion

Here we discuss the results in the context of the three research questions introduced at the end of Section 4.1.

Can vibrotactile or kinesthetic feedback influence VR training performance in a buzz-wire motor skill task?

Improvements (i.e., reductions) in task completion time were seen after overall training (pretest to post-test) as well as Phase 1 training (pretest to intermediate test) for all participants. Similarly, concerning contact time, participants in all feedback conditions showed improvements (i.e., reductions) after overall training (pretest to post-test). Furthermore, statistical tests comparing the improvement scores across the three feedback conditions were inconclusive in finding any difference between them. This implies that, while IVR training with vibrotactile and kinesthetic feedback improves performance, it is not clear which haptic modality is better or if they are better than visual feedback only. However, the analysis points to the potential for further investigation on the use of haptic feedback to modulate arousal and thereby improve performance. Furthermore, alternate uses of haptic feedback could be explored in the context of training in the buzz-wire task, for example in the form of virtual fixtures or guidance (Abbott et al., 2007; Chinello et al., 2017; Devigne et al., 2020; Kuang et al., 2022) .

Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in arousal levels during training? Is there a link between physiological arousal during training and improvements in performance afterward?

Analysis of the data revealed no effect of feedback condition on EDA and HRV. Analysis of SAM responses did not reveal any effect of feedback on subjective arousal. However, pairwise comparisons of EDA and HRV metrics across feedback conditions revealed a higher arousal level (for the HRV metric LF/HF Ratio) among participants receiving kinesthetic feedback compared to those receiving vibrotactile feedback during training. This finding is similar to that of (Krogmeier et al., 2019) where it was observed in a scenario in which participants received haptic feedback (on the torso) from a vest, that the most realistic haptic feedback condition resulted in greater arousal compared to the least realistic. These links between arousal and haptic feedback offer opportunities for further investigation, especially in light of indications from our study that certain arousal levels are correlated with performance improvements. Specifically, correlation tests revealed a weak negative correlation between improvements in performance and two HRV metrics (RMSSD, SDNN), i.e., as arousal increased (decrease in RMSSD and SDNN is correlated with increases in arousal) during VR training, so did performance. This link was further supported by statistical tests which showed higher arousal levels indicated by four HRV metrics (HR, IBI, RMSSD, and SDNN) in the high-improvement group as compared to the arousal levels of participants in the low-improvement group. This trend was again confirmed by statistical tests comparing performance between groups of participants split according to the degree of arousal, which showed that participants in high-arousal groups (defined by the HRV metrics RMSSD and SDNN) demonstrated greater improvement scores compared to those in low-arousal groups. These two findings, that kinesthetic feedback resulted in more arousal (compared to vibrotactile feedback) and that increases in arousal were corelated with increases in performance, inspire future research on the potential for increasing training performance by varying arousal with the help of haptic feedback. Even though our study points to higher levels of arousal linked to better improvements in performance, we also acknowledge the view that increasing arousal or stress may have detrimental effects on performance. For example, in a buzz-wire training scenario, (Radhakrishnan, Chinello, et al., 2022) found that participants across VR and physical training conditions who had the highest arousal levels during training showed the lowest improvements in performance (and vice-versa). Though the motor task is similar (a buzz-wire scenario), that study had participants undergo training in a different setup, i.e., physical training and VR training with simple non-directional vibrotactile feedback from the Oculus Touch controller. There are also non-VR studies that point out that high-arousal levels may lead to lower performance (Prabhu et al., 2010; Quick et al., 2017). (Wu et al., 2010) in their VR driving scenario found that moderate arousal levels (neither too high nor too low) correlated with the best performance. Perhaps, such contradictions may be solved if variables in addition to performance and arousal are considered. For example, a high-arousal level may be linked to higher enjoyment or motivation, which in turn affects performance positively, but similar arousal levels due to anxiety may negatively affect performance.

Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in reported presence, task load, and self-efficacy?

The analysis showed no effect of feedback condition on overall presence. This might be caused by the subjective experience of presence being dominated in this study by the visual aspect which were the same across the three feedback conditions. As (Grassini et al., 2020) had linked increased presence to better training outcomes, the lack of an effect of feedback condition on improvements in performance may thus be partially explained by the lack of differences in terms of presence. However, there are inspirations from literature for further investigation into improving presence and thereby training outcomes. For example, some studies have found vibrotactile feedback coupled with visual feedback to result in better-reported presence compared to visual alone (Gibbs et al., 2022) or kinesthetic feedback alone (Kreimeier et al., 2019).

The analysis also revealed no effect of feedback condition on overall task load. This is important as increases in workload have been linked to decreased motor skill performance as described by (Yurko et al., 2010) in their study on simulator based laparascopy training. In light of this, the lack of differences in performance metrics between the feedback conditions in our study maybe linked to the lack of differences in perceived task load. However, it was also observed that there was a near statistically significant trend towards participants in the visual/kinesthetic condition reporting more temporal demand, i.e., they felt the training to be more rushed/hurried compared to those in the visual/vibrotactile condition. This is interesting as ultimately the outcome of the training in terms of performance were indistinguishable across the feedback conditions in spite of this difference. In a study using a similar buzz-wire test setup, (Radhakrishnan, Chinello, et al., 2022) found participants who received physical training to report more temporal demand than those who underwent IVR training (with non-directional vibrotactile feedback), and similar to our study, overall task load and performance improvements were indistinguishable between the training conditions. On the other hand, (Weber et al., 2013), in a peg-in-hole task using desktop VR, found that vibrotactile feedback resulted in a higher task load compared to kinesthetic feedback, and the authors attributed this to "feedback ambiguity" arising from the design of the vibrotactile feedback. Therefore, it is possible that the lack of differences in overall task load between the kinesthetic and vibrotactile feedback conditions in our study may arise from the lack of directional ambiguity as shown in the perceptual experiment described in Section 4.3.

There were no statistically significant changes in self-efficacy during VR training for all the feedback conditions. This is surprising, as prior literature points to the links between self-efficacy and learning/training performance (Bandura, 1986; Makransky & Petersen, 2021). Our study has shown that there is improvement in performance across all the feedback conditions, in spite of this lack of increase in self-efficacy. (Radhakrishnan, Chinello, et al., 2022) presents a contrasting example with a similar buzz-wire motor skill task, where it was found that significant improvements in self-efficacy during VR training was accompanied by corresponding improvements in performance. (Stevens et al., 2012) in a motor skill training task showed that increased task-difficulty level is one factor which leads to impairments in both self-efficacy and performance. Therefore, future research could consider methods to improve self-efficacy, for example by adapting the difficulty level of training.

Summary

In summary, VR training is effective in improving performance regardless of the considered haptic feedback modality. So, one may wonder if kinesthetic or vibrotactile feedback were to be used, are they interchangeable, or should additional parameters be considered? Our analysis of the data showed that (1) participants in the kinesthetic feedback condition showed higher arousal than those in the vibrotactile feedback condition, and that (2) there is a correlation between higher arousal levels and higher performance across feedback conditions. These findings should encourage further research in using diverse types of haptic feedback to potentially affect arousal levels and performance. Future research could explore variations of kinesthetic feedback, for example with wearable exoskeletons, which might also combine aspects of cutaneous feedback to provide a good trade-off between cost and performance (Pacchierotti, 2022).

4.6 Limitations

Arousal in the present study may be linked to multiple factors including skill level, task complexity, visual feedback, the novelty of VR, and haptic rendering and hardware. Pre-existing skill levels and arousal levels were controlled by using a baseline phase and novelty effects were controlled by a tutorial phase. However, there is scope for further improvement to obtain a more fine-grained link between haptic feedback and physiological arousal. Future studies investigating the effect of haptic feedback on arousal may try to incorporate a baseline haptic feedback condition where the participants receive haptic feedback in regular intervals as seen in (Krogmeier et al., 2019). The corresponding physiological arousal levels can then be measured, for a baseline level of correlation between haptic feedback and arousal to be established.

The Geomagic Touch's workspace is another limiting factor in its use in immersive VR training scenarios. Future studies may explore either wearable haptic devices (Kourtesis et al., 2022; Pacchierotti, 2022; Pacchierotti et al., 2017) or grounded kinesthetic devices with larger workspaces. It is also worth observing that the two haptic feedback conditions have some shared properties. For example, in the kinesthetic condition, there is an element of cutaneous sensation felt while the participants hold the handle and when they receive feedback. Similarly, in the vibrotactile condition, participants feel some resistance due to the inertia inherent in the Geomagic Touch. However, since these properties are shared across the feedback conditions, they have been controlled in this experiment.

Another potential limitation of this study is the between-subjects methodology followed. Though it is better suited for measuring learning effects, this might partly explain the lack of significant findings in subjective metrics. None of the participants in any of the feedback condition could experience any of the other feedbacks available, therefore subjective measures may not reveal the actual effect of the haptic modality on these subjective metrics. One solution is to have a larger sample size in between-subjects studies, which can reveal smaller effect sizes. On the other hand, (Richard et al., 2022), in a simulation study comparing within and between-subject methods for evaluating embodiment, concluded that the within-subjects method is more sensitive in detecting small effect sizes while having smaller sample sizes. However, as within-subjects methodology may potentially cause recency effects (Bjork & Whitten, 1974) affecting the analysis of objective performance measures, future studies might control for this aspect by having longitudinal cross-over studies.

4.7 Conclusion

In this study, a fine motor skill training buzz-wire task in immersive VR was used to investigate the effect of haptic feedback and physiological arousal on performance. The experiment revealed the effectiveness of motor skill training in VR regardless of the haptic feedback. The investigation into physiological arousal levels between three feedback conditions revealed that training with kinesthetic feedback resulted in higher arousal compared to training with vibrotactile feedback. Links between arousal and performance were also found, with increases in arousal being accompanied by increases in performance. The inclusion of haptic feedback thus holds potential for motor skill training in VR, though further research is needed to explore the trade-offs between kinesthetic feedback and varieties of cutaneous feedback, such as the vibrotactile feedback considered here. With regards to investigating the relationship between haptic feedback and arousal, additional physiological metrics can be considered, such as EEG (electroencephalography), EMG (electromyography), and pupil dilation, particularly since biosensors to measure these signals are being increasingly integrated into commercial VR head-mounted displays.

Chapter 5 – A Controlled, Preregistered Experiment on Self-Efficacy and Performance in Adaptive Virtual Training

Immersive virtual reality (IVR) offers novel and promising ways of continuously adapting training difficulty and content to the individual trainee, potentially paving the way for an improved fit between training content and trainee needs. The present paper describes a preregistered, controlled experiment (N = 130) where participants received IVR-based fine motor skill training with a focus on improving speed and accuracy. Using a between-subjects design, participants were randomly assigned to either adaptive training (N = 65), where training content continuously adapted to the behaviour of the trainee, or fixed training (N = 65), where training content was based on a measure of trainee behaviour at the beginning of the study. Results revealed no significant difference between the groups for neither performance nor self-efficacy, suggesting that further research is needed to investigate when the additional complexity of adaptive training is warranted. As for the overall effect of training, participants improved in both accuracy (d = 0.416) and speed (d = 0.580) on a virtual performance test, while performance on a real equivalent (i.e., transfer of skill) showed improved accuracy (d = 0.287) but reduced speed (d = 0.232). The effect of training on measures of self-efficacy were mixed. Results demonstrated that performance measures in IVR should not necessarily be expected to transfer to similar tasks outside IVR, emphasising the need for future studies to include measures of skill transfer when investigating IVR-based training.

5.1. Introduction

5.1.1 Training in IVR

Learning and training in immersive virtual reality (IVR) have been employed across various domains, such as school and university education, rehabilitation, professional training for doctors, and office and industrial workers, focusing on cognitive and psychomotor skills (Jensen & Konradsen, 2018). The literature on training in IVR primarily relates to school and college education (Hamilton et al., 2021) and the teaching of procedural and safety knowledge for industrial training purposes (Feng et al., 2018). In contrast, motor skill training literature in IVR has been dominated by medical use cases, particularly in surgical and dental domains that require fine motor skills (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021). Researchers have investigated the relative advantages of IVR-based training over other media (e.g., video training) and variations within IVR, such as different levels of visual or haptic fidelity (Huber et al., 2018; Jain et al., 2020), user characteristics (Shakur et al., 2015), and training methods (Harvey et al., 2021). However, the results of these studies have been mixed. For example, Pulijala et al. (2018) found IVR to be more effective than presentation training, Hooper, Tsiridis, Feng, Schwarzkopf, Waren, Long, Poultsides, Macaulay, Papagiannakis and Kenanidis (2019) showed IVR to be more effective than physical training for hip arthroplasty surgery, and Butt et al. (2018) observed the same advantage of IVR over physical training for catheter insertion immediately after training, but with no difference after a week. On the other hand, in a comparison of IVR to desktop VR training, Frederiksen et al. (2020) found that IVR was less effective and caused more cognitive load among students of laparoscopic surgery. Thus, the effectiveness of IVR training compared to other types of training remains inconclusive and an open research topic (Checa & Bustillo, 2020), especially in the case of IVR-based motor skill training (Coban et al., 2022).

A fundamental assumption of research on simulation-based training is that skills developed in IVR can be transferred to the real world (Gegenfurtner et al., 2014; Gegenfurtner et al., 2013). However, with prior research highlighting that transfer becomes increasingly difficult the bigger difference there is between the context during training and the context when retrieving the skill later (Ragan et al., 2015) it is no surprise that skill transfer of IVR-based training has been questioned (Jensen & Konradsen, 2018). Although there is support for virtual training having a beneficial effect outside the virtual environment (Cooper et al., 2021; Murcia-Lopez & Steed, 2018), it is not uncommon for research to purely include IVR-based outcome measures without a measure of skill transfer (Frederiksen et al., 2020; Huber et al., 2018; Lang et al., 2018). Including measures of performance both in- and outside IVR would provide valuable information about the potential loss that is to be expected when transferring a skill from one setting to another.

5.1.2 Self-efficacy

The term self-efficacy is used to refer to one's perceived capabilities for learning of performing an action (Schunk & DiBenedetto, 2016). Research has supported the role of self-efficacy not only in relation to learning outcomes, but also in training (Gegenfurtner et al., 2014). Specifically, self-efficacy is an important predictor of performance on the task at hand (Feltz et al., 2008; Moritz et al., 2000; Rosenqvist & Skans, 2015) as well as future performance on similar tasks (Pascua et al., 2015; Stevens et al., 2012). For example, when Chauvel et al. (2015) had participants practise golf putting on holes made to be perceived as either big or small, self-efficacy was found to be significantly higher when the hole was perceived as larger (i.e., making the task seem easier). Furthermore, although the holes were the same size, performance was significantly higher for the task perceived as easier, both on the task and on a retention task the following day. The same result has been replicated in multiple tasks related to motor skills, such as golfing (Abbas & North, 2018; Chauvel et al., 2015), dart throwing (Ong et al., 2015), and soccer (Mousavi & Iwatsuki, 2021). When developing training content, it is therefore crucial to consider how to increase training outcome by supporting the individual's level of self-efficacy.

According to the social cognitive theory, major influences of self-efficacy include personal experiences of success, vicarious experiences (i.e., observing others succeed), social persuasion, and physiological factors (Bandura, 1977). Amongst these, the most important factor is personal experiences of success (Schunk & DiBenedetto, 2016). in other words, the experiences of being able to complete a task or accomplish a goal at a satisfactory level is a key predictor of the individual's level of self-efficacy, which in turn predicts performance gain. As such, the majority of research on self-efficacy and skills training has focused on the role of perceived success (Wulf & Lewthwaite, 2016), supporting a stronger link between performance feedback and improvements when the feedback is given for successful rather than unsuccessful trials (Abbas & North, 2018; Saemi et al., 2012) as well as improved retention and skill transfer (Wulf et al., 2014).

According to the cognitive affective model of immersive learning (CAMIL), IVR is characterized by higher levels of presence and agency than traditional media, which in turn supports learning outcomes through cognitive and affective factors such as self-efficacy (Makransky & Petersen, 2021). Indeed, when compared with less immersive media, IVR tends to show beneficial effects on self-efficacy in an educational setting (Huang et al., 2022; Klingenberg et al., 2020; Shu et al., 2019). However, while multiple studies have found evidence of the role of self-efficacy in simulation-based training (Gegenfurtner et al., 2014; Gegenfurtner et al., 2013), only few have narrowed the scope to IVR-based training (Buttussi & Chittaro, 2017; Lehikko, 2021; Liu et al., 2022; Pulijala et al., 2018; Radhakrishnan, Chinello, et al., 2022; Song et al., 2021). Furthermore, these rarely include sufficient measures of self-efficacy and training outcomes to fully explore the role of self-efficacy in relation to the effectiveness of the training.

Thus, an abundance of research supports the link between self-efficacy and learning or training as well as the role of feedback during training. Given the role of self-efficacy in relation to performance gain, it is expected to be especially relevant for IVR-based training, which has the potential to be formed and tailored to the individual in ways impossible with prior technology. That said, limited research has investigated ways to utilise the role of trainee self-efficacy when designing IVR-based training content. The present study aimed to address this gap by investigating options for implementing and utilising dynamic measures of self-efficacy throughout training with the aim of increasing both trainee self-efficacy and performance.

5.1.3 Adaptive training

Tailoring, or adapting, training content to the individual trainee is not a new approach. On the contrary, it has been over 50 years since Kelley (1969, p. 547) described adaptive training as requiring that "*[…] performance be continuously or repetitively measured in some way, and that the measurement be employed to make appropriate changes in the stimulus, problem, or task.*" In other words, in adaptive training technology is used in the place of a skilled instructor with the aim of continuously monitoring the responses of the individual and adjusting training content for optimal training outcome. Non-adaptive training, where a one-size-fits-all approach is used, is generally easier and cheaper to implement, but with the risk of a mismatch between training content and trainee needs in relation to factors such as difficulty, engagement, and training focus (Zahabi & Abdul Razak, 2020).

A further distinction is made between *fixed training*, where training content is adjusted exclusively based on measures prior to the start of the training session, and *adaptive training*, where measures of trainee behaviour and adjustment of training content happens throughout the training session (Gerbaud et al., 2009; Kelley, 1969). While fixed training is relatively easy to implement, it fails to take into account the difference in improvement rates between individuals, resulting in the training content being too easy for some individuals and too challenging for others (Kelley, 1969). Therefore, aiming to implement adaptive training is to be preferred, but requires three core elements. First, a continuous measure of trainee behaviour is required to monitor relevant aspects of the individual's changing performance throughout the session (Kelley, 1969). The term *trainee behaviour* is used here to highlight that the measure need not relate to performance on the task at hand, but can also be measures of learning style, psychophysiological measures, such as eye-tracking, ECG, or EDA, or psychological factors, such as self-efficacy, immersion, or engagement. Furthermore, the chosen behaviour can either be measured throughout the whole training session or between different parts of the training session. As for the second core element, a feature of the training content, termed *adaptive variable*, must be chosen based on relevance for training outcome, such as training difficulty, feedback, or training focus (Zahabi & Abdul Razak, 2020). Third, *adaptive logic* is implemented to describe the relationship between the trainee behaviour and the adaptive variable, such as increasing the difficulty of training when trainee performance increases.

It is generally assumed that the increased complexity of implementing fully adaptive training (as opposed to fixed training) is accompanied by increased training outcome through the superior fit between trainee needs and training content (Kelley, 1969; Zahabi & Abdul Razak, 2020). In a systematic review of current usage of virtual reality-based adaptive training, Zahabi and Abdul Razak (2020) notes that most studies were concept or feasibility studies and as such did not investigate the effectiveness of adaptive training. Of the few that did, the results are mixed, with some support for adaptive over non-adaptive training (Lang et al., 2018; Ma & Bechkoum, 2008; Wang et al., 2017), whereas other studies found no difference between the two (Billings, 2012; Serge et al., 2013).

Although this indicates that the use of adaptive training is already being investigated, only few studies utilised truly adaptive training (as opposed to fixed training; e.g., Mariani et al., 2020) and even fewer has done so using *immersive* virtual reality (Vaughan et al., 2016; Zahabi & Abdul Razak, 2020). The few current explorations of IVR-based adaptive training were mainly based on concept development (e.g., de Lima et al., 2022; Drey et al., 2020; Iván Aguilar Reyes et al., 2022) and pilot studies (e.g., Chiossi et al., 2022; Muñoz et al., 2016). Therefore, robust, controlled experiments with larger samples are required to move beyond concept development of adaptive IVR training, with the aim of investigating the effect of adding the additional layer of complexity as compared to fixed training. While doing so, research should ideally also implement measures of transfer of skills from virtual training to the desired setting to investigate the effect of training features on skill transfer. Furthermore, Zahabi and Abdul Razak (2020) highlights the need for adaptive training where trainee kinematic or kinetic information is used for continuous adaptation content and where the used adaptive variable goes beyond content difficulty, for example by providing adaptive feedback to the user.

To address these gaps, the present study aimed to compare adaptive and fixed IVR-based training in a controlled experiment, following the assumption of Kelley (1969) that higher levels of adaptiveness should be superior for training outcome. Moreover, following the arguments of Zahabi and Abdul Razak (2020), adaptive logic was formed around multiple sources of information, including performance data and psychological measures of self-efficacy. Lastly, adaptive variables were based on not only adjusting task difficulty, but also features of training content in the form of training focus.

5.1.4 Hypotheses

The primary aim of the present study was thus to investigate the use of adaptive IVR-based training in a controlled experiment. Two hypotheses were formulated in relation to the effect of adaptive IVR training whereas the third focused on the relationship between self-efficacy and performance.

- H1: The majority of prior research has made the case that adaptive training should be superior to fixed training (Kelley, 1969; Zahabi & Abdul Razak, 2020). Therefore, adaptive IVR training was expected to lead to higher performance as compared to fixed adaptive IVR. To investigate differences in skill transfer, performance (in for the form of increased accuracy and speed) were measured both in IVR and on a physical version of the same task.
- H2: Secondly, research on self-efficacy suggests that adjusting difficulty and content to make successful experiences possible and clear to the trainee is of great importance when aiming to support trainee confidence (Hutchinson et al., 2008; Saemi et al., 2012). Based on the increased ability to adapt to the needs of the individual trainee, adaptive training was expected to lead to higher self-efficacy than fixed training.
- H3: A third hypothesis focused on the relationship between self-efficacy and performance, where self-efficacy was expected to affect the relationship between training and performance.

Additionally, exploratory analyses focused on investigating the level of transfer from virtual training to a real equivalence of the task as well as dynamic changes in performance and self-efficacy throughout virtual training.

5.2 Study

5.2.1 Method

The study was conducted at Cognition and Behavior (COBE) Lab at Aarhus University and preregistered using AsPredicted. Preregistration, study materials, and study data are available on the Open Science Framework (OSF; https://osf.io/jmg9t/).

See Figure 49 for an overview of the experiment flow. After arriving at the lab, participants completed one repetition of the physical version of the buzz wire task as a familiarization task and a short demographic questionnaire. Before and after IVR training, a physical and virtual version of the buzz wire task were completed to measure changes in performance. Before the two physical versions of the task, self-reported self-efficacy were measured using a 6-item scale. During the virtual training, participants were randomly assigned to one of two conditions, adaptive or fixed training, and completed 10 repetitions of the virtual buzz wire task with changing training focus and difficulty. After each repetition of the virtual wire, participants rated their confidence in completing the task *quickly* and *accurately* to measure dynamic changes in self-efficacy, which was also used to adjust adaptive logic for the training. The study duration was approximately 35 minutes and participants were paid 85DKK (approximately 11.5 euro) for taking part in the study.

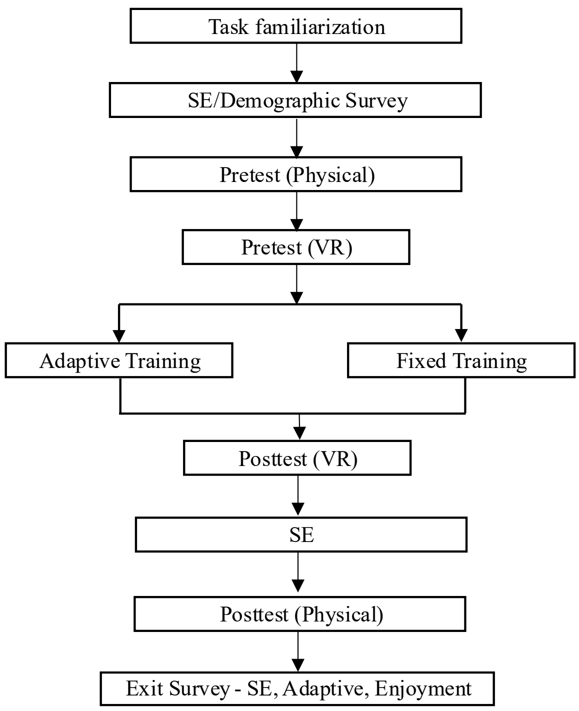


Figure 49 Flow of experiment

5.2.1.1 Sample

The final sample consisted of 130 individuals (Mage = 25.37, SD = 6.07; 51.5% female). The sample size was determined based on a priori power analysis to estimate the required sample size to detect a medium effect in the primary analysis using one-tailed tests with high statistical power ($1 - \beta = 0.80$) and an alpha level of 0.05. All statistical analyses include the full sample of N = 130 unless stated otherwise.

5.2.1.2 Motor skill task

(a) (b)

Figure 50 Buzz-wire test in the (a) VR setup (b) Physical setup

In the study, participants were instructed to grasp the handle as illustrated in Figure 50 and move the connected metal loop along a wire as quickly and accurately as possible, minimizing contact between the loop and the wire. The handle in the physical version was a 3D printed replica of the Oculus Quest controller, while the controller itself was used in the IVR task. Both the virtual and physical tests featured wires measuring 52 cm in length with eleven 90° bends spanning the x, y, and z axes from the starting point (labelled 'A') to the finishing point (labelled 'B'). For the physical setup, electrical circuits detected contact between the loop and the wire, as well as between the loop and the associated start (‘A’) and end points (‘B’). This data was transmitted to the iMotions data collection platform.

The IVR test setup was created with the Unity3D game engine and designed to maintain identical wire dimensions to the physical test. Collision detection code implemented using C# scripting inside Unity3D measured the contact between the virtual loop and the virtual wire. The data was then used to measure total contact time, defined as the time when the loop is in contact with the wire (in seconds), as well as the task completion time, defined as the time taken to move the loop from A to B (in seconds).

.

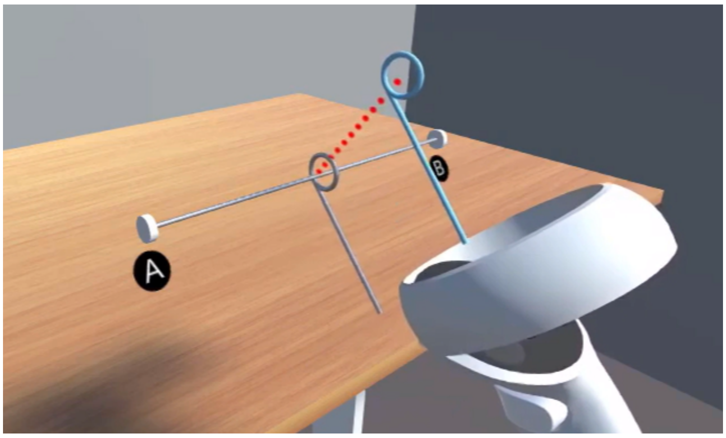


Figure 51 Ghost effect when the loop is moved out of the wire in the VR setup.

When the participant made a mistake in the virtual setup, i.e., the loop touched the wire, there was nothing to physically restrict the participant's hand, unlike the physical version where there was an actual wire to provide resistance. To provide a feeling on the hand of the participant when the loop touched the wire, a haptic vibration was provided using the Oculus controller. However, by the time contact is made, the loop would already have passed through the wire creating an unrealistic effect for the participant. To solve this, a 'ghost effect' (Figure 51) was programmed to show a blue translucent loop at the contact position where the loop passes through the virtual wire. A dotted red line indicated the direction where the user should move to re-join the wire, thus helping the participant understand how to bring their loop back into the wire, at which point the blue translucent 'ghost' disappears.

5.2.1.3 Virtual Training

Participants utilized an Oculus Quest 2 head-mounted display (HMD) connected to a PC, functioning in PC VR mode. The virtual environment was developed with the Unity3D (version 2019.4) game engine. The Oculus SDK supplied the position and rotation information for both the controller and the HMD, which were subsequently applied to control the virtual loop and the participant's viewpoint in the three-dimensional space of the virtual setting (see Figure 52). To reduce novelty effects, participants initially performed a virtual task that required moving the loop along a short, straight wire, familiarising themselves with the mechanics of IVR before starting the main training. The training consisted of ten trials that focused on either speed or accuracy. The wire was 57 cm long (from end to end), featuring eighteen 90-degree bends and extending 21 cm horizontally. This configuration was maintained across all ten trials.

.

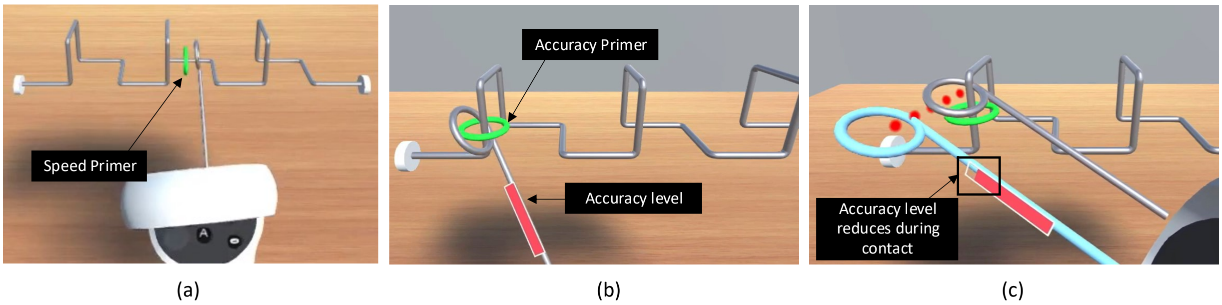


Figure 52 Types of VR training (a) speed focused training, (b) & (c) accuracy focused training

Speed focused training: A green speed primer ring moved at a constant pace, determined by the participant's prior performance. Participants were instructed to concentrate on completing the task by matching or surpassing the speed primer's pace while maintaining the requisite level of accuracy.

Accuracy focused training: An accuracy primer, represented by a green ring, moved in tandem with the participant's loop, maintaining optimal orientation and distance from the wire, serving as a reference to minimize errors. A red bar signified the accuracy level, where 100% corresponds to the accuracy from a previous trial. The bar diminished in size as the participant made contact, proportionate to the current total contact time and the previous total contact time. For example, if the participant had a total contact time of 5 seconds in the previous trial and the current contact time reaches 2.5 seconds, the red bar would shrink by 50%.

5.2.1.4 Adaptive logic and conditions

Two sets of adaptive logic were developed for the study. The first was based on the majority of prior research (Iván Aguilar Reyes et al., 2022; Zahabi & Abdul Razak, 2020), using performance measures (trainee behaviour) to adjust difficulty of the primer for each repetition (adaptive variable). The second adaptive logic used participant self-efficacy (trainee behaviour) to adjust the training focus to either speed- or accuracy focused training (adaptive variable).

In the beginning of the study, participants were randomly assigned to fixed or adaptive training. In fixed training, training content was based on performance and self-efficacy during the initial virtual pre-test, whereas training content in adaptive training was continuously adjusted based on performance and self-efficacy in the most recent repetition of the task (see also Figure 53).

With fixed training, performance and self-efficacy were measured after the virtual pre-test, whereafter difficulty was set to match the performance of the participant with a fixed increase in difficulty of 3% each repetition. The exact increase in difficulty was based on prior research (Radhakrishnan et al., 2023), matching the average improvements in performance on the virtual buzz wire task. Additionally, fixed training included 8 out of 10 repetitions of the type of training (speed vs. accuracy) each participant reported lowest confidence in.

With adaptive training, the difficulty of each repetition of the task was set to match the performance of the participant in the last repetition of the same type. Training focus was continuously adjusted in the same way, by including the type of training that was reported the least amount of confidence in during the last repetition. To ensure that no participant received only one type of training focus, the adaptive training would include no more than 8 of the same type of training focus.

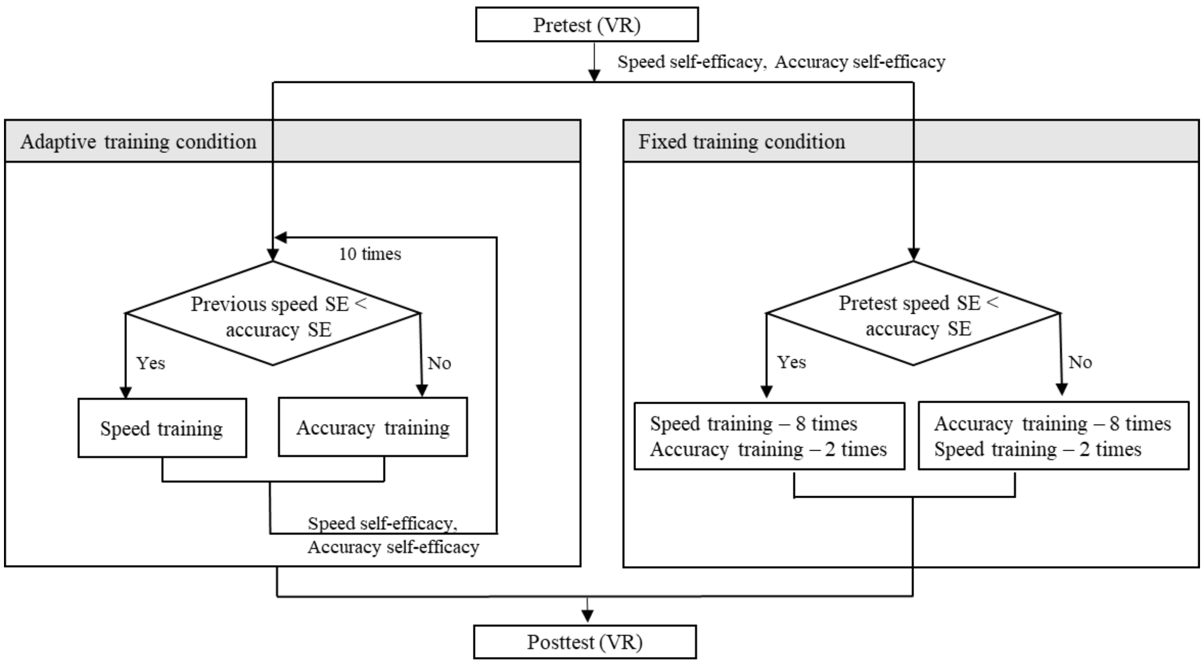


Figure 53 Adaptive logic for the two training conditions.

5.2.1.5 Self-efficacy

A scale were constructed based on Bandura (2006) to measure self-efficacy. Participants were asked to rate their confidence by typing a number from 0 (*Not at all certain*) to 100 (*Highly certain*) on six items. The items related to completing the buzz wire task with high accuracy (e.g., *I can complete the buzz wire task with a minimal number of touches*), high speed (e.g., *I can control the fine motor movements necessary to complete the buzz wire task quickly*), or a combination of the two (e.g., *I can maintain a high level of speed and accuracy while completing the buzz wire task*). The full instructions and remaining items can be found on OSF. As described in the preregistration, main analyses were based on change in self-efficacy on the overall SE, calculated by subtracting the pre-measured SE (M = 54.09, SD = 19.31) from the post-measured SE (M = 45.31, SD = 19.07). All items were highly correlated (correlations between r = .577 and .854) with sufficiently high Cronbach's alpha (α = .941; α = .943).

After the virtual pre- and post-test as well as after each repetition of the buzz wire, participants were asked to rate their confidence in completing the buzz wire task *quickly* and *accurately* on a scale from 0-100. These items were used to inform the adaptive logic and for explorative analyses focusing on dynamic changes in self-efficacy.

5.2.1.6 Performance

Prior studies using the same buzz wire task measured performance on one repetition of the buzz wire task before and after training (Radhakrishnan, Chinello, et al., 2022; Radhakrishnan et al., 2023). To increase sensitivity of the outcome measure, the present study lengthened the pre- and post-measure by having participants complete the task in both directions.

Two performance measures were used based on prior research on motor skills training (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021; Radhakrishnan et al., 2023): speed and accuracy. Speed was measured as task completion time, meaning the time the participant spent on moving the loop through the wire. Accuracy was measured as contact time with the wire, indicating how much a participant touched the wire during the task. To convert these measures to a meaningful measure of improvements in performance and to account for differences in baseline performance, the main performance measures were the decrease in task completion time (i.e., improved speed) and decrease in contact time (i.e., improvement in accuracy) measured in percentage.

Additionally, the majority of prior literature investigating IVR-based training measured improvements in performance either inside IVR (Lang et al., 2018) or outside IVR (e.g., Koumaditis et al., 2020a; Murcia-Lopez & Steed, 2018), but very rarely both (e.g., Sportillo et al., 2015). In the present study, performance was measured both on a virtual and physical version of the buzz wire task to compare improvements between the two, thus reflecting skill transfer.

5.3 Results

5.3.1 Main effect of virtual training

First, paired t-tests were conducted to investigate the overall effect of virtual training on performance and self-efficacy. Analysis of the main effect of training revealed that participants completed the virtual wire in a significantly shorter amount of time (pre-test M = 52.333, SD = 19.220; post-test M = 44.330, SD = 17.802), t (129) = -6.612, p < .001, 95% CI [-10.398, -5.609], d = -0.580, and with significantly lower contact time (pre-test M = 6.709, SD = 4.961; post-test M = 5.06, SD = 2.739), t (129) = -4.748, p < .001, 95% CI [-2.335, -0.962], d = -0.416. On the physical wire (i.e., skill transfer) performance increased for contact time (pre-test M = 15.975, SD = 5.572; post-test M = 14.462, SD = 6.376), t (129) = -3.277, p < .001, 95% CI [-2.426, -0.599], d = -0.287, whereas training resulted in a significant slower completion time (pre-test M = 53.278, SD = 24.093; post-test M = 57.526, SD = 25.790), t (129) = 2.644, p = .009, 95% CI [1.070, 7.426], d = 0.232. In short, training led to participants completing a virtual wire both quicker and with higher accuracy while completing a real wire slower but with higher accuracy than prior to training. Furthermore, virtual training had a medium effect on the virtual wire (d = -0.580 for speed & d = -0.416 for accuracy) but only a small effect on the physical wire (d = -0.287 & d = 0.232), indicating that although skill transfer was observed, it was in a relatively limited degree. Interestingly, improved accuracy on the virtual wire was positively associated with improved accuracy (r = .201, p = .022), but not speed (r = -.059, p = .503), on the physical wire, whereas improved speed on the virtual wire was positively associated with improved speed (r = .456, p < .001), but not accuracy (r = .075, p = .399), on the physical wire (see also Table 19).

Unexpectedly, self-efficacy was significantly lower after training (M = 45.309, SD = 19.073) as compared with before training (M = 54.091, SD = 19.305), t (129) = 5.345, p < .001, 95% CI [5.531, 12.033], d = 0.469. To further explore trainee confidence, two repeated measures ANOVA were conducted, analysing dynamic changes in self-efficacy for accuracy and speed throughout the 10 training repetitions. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated for both accuracy, $\chi^2(44) = 149.79$, $p < .001$, and speed, $\chi^2(44) = 141.45$, $p < .001$, therefore the Greenhouse-Geissor correction tests were used. Tests of within-subjects effect showed a significant main effect of time for both self-efficacy for accuracy, $F\ (6.78, 873.8) = 12.856$, $p < .001$ partial $\eta^2 = .091$, and speed, $F\ (7.08, 1161) = 13.997$, $p < .001$, partial $\eta^2 = .098$. Furthermore, a significant linear trend was found for self-efficacy over time for both accuracy, $F\ (1, 129) = 32.944$, $p < .001$, partial $\eta^2 = .203$, and speed, $F\ (1, 129) = 18.445$, $p < .001$, partial $\eta^2 = .125$, indicating a general increase in self-efficacy for both accuracy and speed when measured after each training repetition (see also Figures 54 and 55).

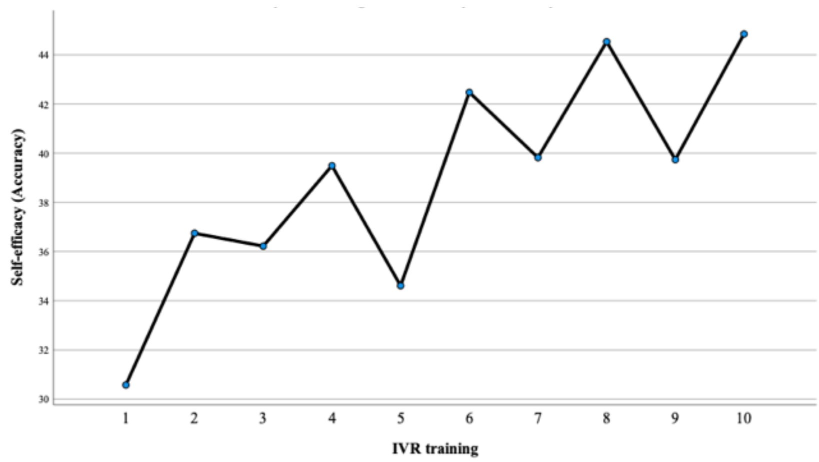


Figure 54 Dynamic change in self-efficacy for accuracy throughout virtual training

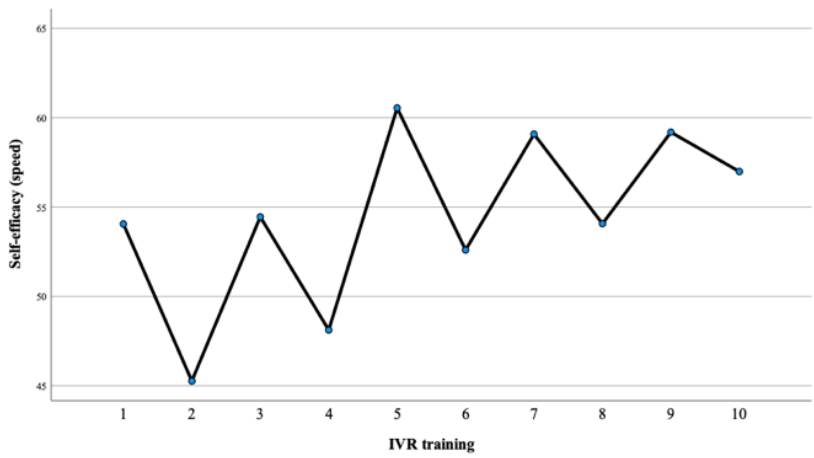


Figure 55 Dynamic change in self-efficacy for speed throughout virtual training

5.3.2 Main analysis: Adaptive and fixed training

As specified in the preregistration, one-sided t-tests were utilised to investigate hypothesis 1 and 2, comparing adaptive (N = 65) and fixed (N = 65) virtual training. One-sided tests were used to increase power of the main analysis, justified by the fact that the study was preregistered and hypotheses where directional (Cho & Abe, 2013; Ruxton & Neuhäuser, 2010). The groups did not differ in age (*p* = .841), or any relevant baseline measures such as self-efficacy (*p* = .308) or aspects of performance (*p* = .2 – *p* = .42). Although not significantly different, the adaptive group consisted of slightly more male participants (57%) than the fixed group (40%; *p* = .054).

H1 assumed that performance would be higher as a result of adaptive training. This was not supported by the results, where no significant difference were found between the groups for completion time (IVR; *t* (128) = -0.018, *p* = .493, 95% CI [-11.724, 11.516], *d* = .0.003), contact time (VR; *t* (128) = 0.339, *p* = .368, 95% CI [-20.834, 29.442], *d* = 0.059) transfer completion time (*t* (128) = 0.00, *p* = .5, 95% CI [-9.776, 9.777], *d* = .000), or transfer contact time (*t* (128) = 0.261, *p* = 397, 95% CI [-10.659, 13.903], *d* = 0.046; Figure 56). Exploratory analysis revealed no significant difference in training difficulty between the groups for neither accuracy focused training, *t* (128) = 0.034, *p* = .973, 95% CI [-1.155, 1.194], *d* = 0.006, or speed focused training, *t* (128) = 1.807, *p* = .073, 95% CI [-0.279, 0.614], *d* = 0.317.

H2 assumed that adaptive training would lead to a higher increase in self-efficacy than fixed training, which was also not supported by the results, revealing no statistically significant differences in changes in self-efficacy (*t* (128) = -0.546, *p* = .293, 95% CI [-8.320, 4.720], *d* = -0.096).

Table 18 Group statistics for adaptive and fixed training

	Condition	N	Mean	Std. Deviation
Self-efficacy change	Adaptive training	65	-9.6821	18.15258
	Fixed training	65	-7.8821	19.39463
Improved accuracy	Adaptive training	65	42.7925	75.56506
	Fixed training	65	38.4887	69.14583
Improved speed	Adaptive training	65	23.1845	34.94697
	Fixed training	65	23.2883	31.94336
Improved accuracy (Transfer)	Adaptive training	65	19.2145	37.31522
	Fixed training	65	17.5926	33.33824
Improved speed (Transfer)	Adaptive training	65	-4.0281	25.97489
	Fixed training	65	-4.0284	30.20296

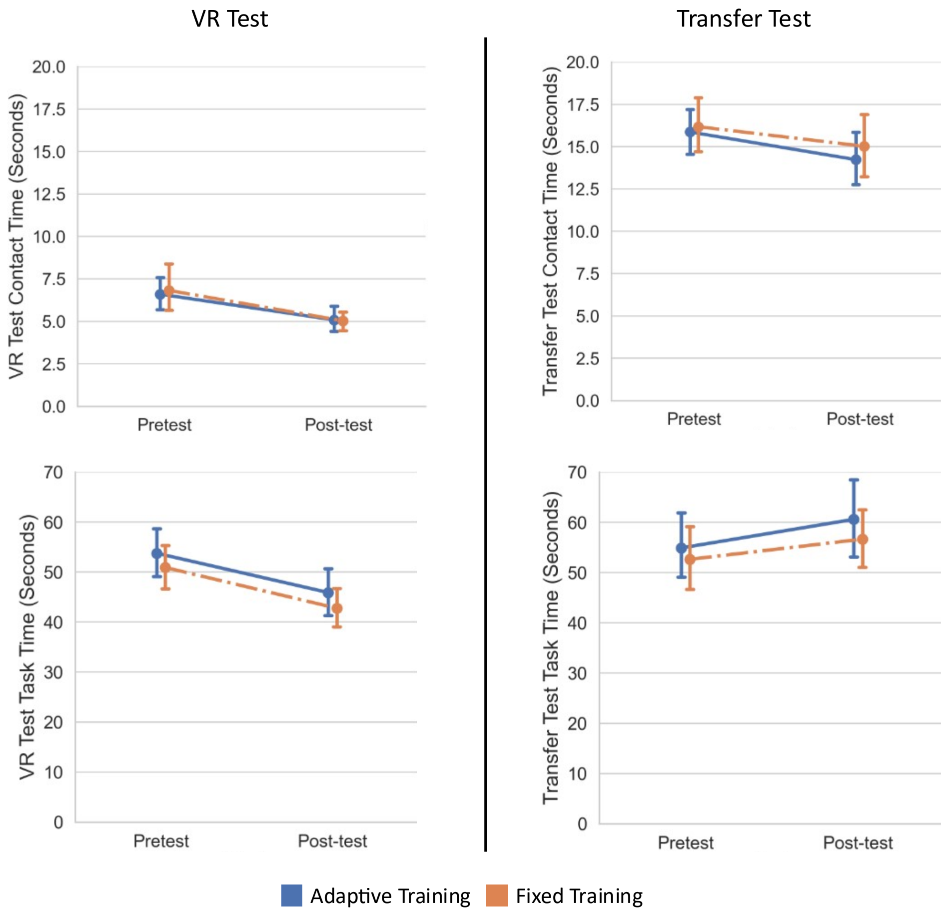


Figure 56 Performance change for both VR and transfer tests for adaptive and fixed training

5.3.3 Exploratory analysis: training type and training success

Training type

During virtual training, each repetition of the buzz wire focused on either speed or accuracy, depending on the individual's level of self-efficacy in the previous repetition (adaptive training) or in the virtual pre-test (fixed training). On average, participants received accuracy-focused training in 6.5 out of 10 repetitions of the buzz wire task, with the adaptive group (M = 7.062, SD = 1.144) receiving significantly more accuracy-focused training than the fixed group (M = 5.877, SD = 2.890), t (128) = 3.072, p = .003, 95% CI [0.422, 1.948], d = 0.539. On average, the amount of accuracy-focused tasks received by participants were negatively correlated with experiences of success (i.e., the number of repetitions where the participant outperformed the primer; r = -.431, p < .001) and change in self-efficacy (r = -.239, p = .006), suggesting that more accuracy-focused training (as opposed to speed-focused training) were more difficult and associated with a drop in self-efficacy.

Training success

As described in the introduction, prior experiences of success are a key predictor of self-efficacy. Therefore, additional analyses were conducted to explore the role of outperforming the primer during training. On average, participants performed better than the primer on 6.2 out of 10 repetitions of the task (M = 6.192, SD = 2.179) with no significant difference between the two experimental groups, t (128) = -0.683, p = .496, 95% CI [-1.019, 0.496], d = -0.120. Furthermore, experiences of success (i.e., outperforming the primer) was positively associated with self-efficacy and improvements in performance inside but not outside IVR (see Table 19). Thus, frequently outperforming the primer was associated with gaining more self-efficacy from the virtual training as well as showing stronger improvements on a virtual wire, but not on a physical wire.

Table 19 Correlation matrix of experiences of success during training, self-efficacy, and performance (N = 130)

Correlations

		Experiences of success during training	Self-efficacy change	Improved speed	Improved accuracy	Improved accuracy (Transfer)	Improved speed (Transfer)
Experiences of success during training	Pearson Correlation	--					
	N	130					
Self-efficacy change	Pearson Correlation	.238**	--				
	Sig. (2-tailed)	.006					
	N	130	130				
Improved speed	Pearson Correlation	.263**	-.117	--			
	Sig. (2-tailed)	.003	.184				
	N	130	130	130			
Improved accuracy	Pearson Correlation	.232**	.108	.066	--		
	Sig. (2-tailed)	.008	.221	.456			
	N	130	130	130	130		
Improved accuracy (Transfer)	Pearson Correlation	.035	.147	.075	.201*	--	
	Sig. (2-tailed)	.696	.096	.399	.022		
	N	130	130	130	130	130	
Improved speed (Transfer)	Pearson Correlation	-.026	.078	.456**	-.059	.459**	--
	Sig. (2-tailed)	.773	.379	<.001	.503	<.001	
	N	130	130	130	130	130	130

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

5.4 Discussion

Adaptive virtual training

Kelley (1969) suggested that when done correctly, adaptive training should be more effective than fixed or non-adaptive training. However, results of the present study revealed no difference in terms of self-efficacy or improved performance either in- or outside IVR. Although contrary to the argument by Kelley (1969), the present result is similar to prior research on adaptive training, which has often found no significant difference between adaptive and fixed training (Zahabi & Abdul Razak, 2020). Two interpretations of this finding are that the difference between groups were either 1) too small or 2) not meaningful when aiming to enhance self-efficacy and motor skills, each of which will be considered below.

First, the lack of difference could be due to training content being too similar. Prior research has shown mixed results when comparing adaptive training to non-adaptive (one-size-fits-all) training. However, the present study compared adaptive training with fixed training, thus further reducing the difference between conditions. Specifically, both conditions included a degree of adaptability, meaning that the lack of difference between groups could be due to the increased adaptiveness being unnecessary. In other words, the fundamental adaptability of fixed training, where a baseline performance and self-efficacy was used to form the focus and difficulty of the training content, may be enough to ensure a good fit of training content without the need for the complex adaptive logic of the adaptive training. This interpretation is further supported by the fact that no differences were found in training difficulty between groups, suggesting that the added adaptiveness from adaptive training were not required to achieve a sufficient fit between trainee need and training content.

Second, the lack of difference could also be a result of choices of irrelevant trainee behaviours, adaptive variables, or adaptive logic. When designing the adaptive logic, two adaptive variables (task focus and difficulty) and two trainee behaviour (performance and self-efficacy) were used to adjust training content to the individual trainee. In terms of adaptive logic, participants received training focusing on the type of performance they felt the least confident in with the aim of improving confidence for said aspect of their own ability. Although this resulted in the adaptive training including significantly more accuracy-focused training, both groups primarily received accuracy focused training in a relatively similar sequence. Interestingly, accuracy-focused training was generally more difficult, indicated by a negative correlation with training success. Furthermore, accuracy-focused feedback was provided by giving participants information every time they touched the wire (i.e., made a mistake) and only indirectly highlighted success in the form of completing the task without letting the red bar run out. Where prior research emphasizes the importance of giving feedback during successful training (Abbas & North, 2018; Saemi et al., 2012), the accuracy-focused training could have provided a too strong emphasis on making mistakes (the red bar getting smaller) at the cost of an emphasis of success (completing without letting the bar run out). A solution for similar future research is to either flip the adaptive logic, thus allowing participants to receive the type of training they feel the most confident or successful in, or adjusting training content to include a clear focus on successfully improving in the type of task they feel the least confident in.

Generally, further research into appropriate selection of trainee behaviour and adaptive variables is needed to better understand their impact on training outcomes as well as their potential for adaptive training. Where the present study focused on performance and self-efficacy as trainee behaviour, a fruitful direction for future research is the inclusion of feedback type and the result of adaptive feedback on self-efficacy. Since novel IVR-based technology are increasingly implementing eye-tracking or psychophysiological measures (ECG, EDA, etc.), another direction is to focus on the potential of utilizing these for trainee behaviour. Lastly, the present study highlights the need for more controlled experiments with relatively large samples to further investigate when the higher requirements of implementing adaptive, rather than fixed, training can be expected to improve training outcomes.

Virtual training, performance, and skill transfer

Virtual training had a significant effect on performance both in- and outside IVR. Virtual training had a stronger effect on performance within the same virtual environment, as indicated by a medium effect size for performance on the virtual task and only a small effect size for performance on the physical task. In accordance with prior research (Jensen & Konradsen, 2018; Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021), this suggests that, although virtual training has an effect on performance outside IVR, the transfer of skill from the virtual environment to the real environment should be expected to come at a cost.

Additionally, on the measure of skill transfer virtual training only led to an increase in accuracy whereas the opposite was true for speed, which was decreased significantly as a result of training. In other words, for one measure of performance, virtual training had the opposite effect on the virtual task as the real equivalent. This finding is in line with the study by Ragan et al. (2015), suggesting that performance in a simulation should not uncritically be assumed to reflect performance on a real version of the same task. Therefore, future research should be mindful of how performance is measured and whether it can be assumed to transfer from a virtual to a real environment. Like in the present study, including measures of the desired outcome variable in both settings offer great potential in terms of not only investigating skill transfer, but also gaining knowledge of the potential loss of performance when seeking to transfer the learned skill to the real world.

The role of self-efficacy in training

The present study provides valuable insight into the role of self-efficacy in IVR-based training. Previous literature supports the central role of self-efficacy in predicting learning (Sitzmann & Ely, 2011) and training outcomes (Gegenfurtner et al., 2014). However, while virtual training did increase performance in the present study, the opposite was true for self-efficacy, which was significantly lower after the virtual training than before.

As discussed earlier, it is possible that feedback of the accuracy-focused training had too much emphasis on highlighting mistakes (i.e., failure) rather than increased accuracy (i.e., success). As supported by prior research, it is crucial that feedback focus on training success, which is associated with higher increases in self-efficacy (Abbas & North, 2018; Saemi et al., 2012). This interpretation is further supported by the fact that receiving more accuracy-focused training was negatively associated with self-efficacy in the present study. Thus, further research should carefully consider ways to adjust training content while also controlling feedback type with the aim of supporting trainee self-efficacy.

Alternatively, the decrease in self-efficacy could be a result of participants overestimating their own ability during the initial measure. As suggested by Bandura (2006), self-efficacy was measured before the behaviour of interest (i.e., completing the real buzz wire task). To give a frame of reference when filling the SE items without receiving training on the task, participants were asked to simply move the loop from one end of the wire to the other at the very beginning of the experiment. However, one explanation of the decrease in self-efficacy is that this familiarization task led to an overestimation of one's own skill whereafter the real task was experienced as significantly more difficult. This would also offer some explanation to the results of exploratory analyses of dynamic changes in self-efficacy, showing a general increase in self-efficacy for both speed and accuracy when measured throughout the virtual training. The contradictory finding emphasises the need for considering multiple measures of concepts such as self-efficacy, as well as including dynamic measures when possible.

5.5 Limitations

One limitation of this study is the short-term nature of the training. The distinction between training and practice effects is crucial in understanding the effectiveness of adaptive training. The relatively short-term training may have led to practice effects rather than genuine training effects (Magill and Anderson). Longer training or training over an extended period may lead to different results. Additionally, this study did not include a retention test, e.g., after an interval of one or two weeks as found in other IVR training literature (Carlson et al., 2015; Murcia-Lopez & Steed, 2018). Though prior literature has found retention tests to reflect positive performance improvements observed from immediate tests (Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, et al., 2021), there have been exceptions as well (Carlson et al., 2015; Meyer et al., 2019). Sometimes, even regressions in performance from immediate tests may be observed (Magill & Anderson, 2016), which could be a natural part of the skill acquisition process. Thus, retention tests after longer intervals are essential.

Another limitation is the study's focus on transfer from one setting (IVR) to another (physical), while the transfer of a motor skill to a completely different task using the same motor skill remains unexplored (Magill & Anderson, 2018b). Considering various types of transfer is crucial, as the loss in performance could be larger when not only changing the setting but also the characteristics of the task (Magill & Anderson, 2018a). Moreover, the experimental design may have led to participants across conditions focusing too strongly on mistakes rather than successes, which could have influenced the outcomes of the study. Finally, there might be an issue with the self-efficacy (SE) measure. Participants may have overestimated their confidence in the initial test, especially in the in-VR SE measures, where they reported higher SE at the beginning. This overconfidence might have resulted from the familiarization task, where participants completed the task without speed or accuracy constraints. Future research should address these limitations to provide a more comprehensive understanding of the effectiveness of adaptive training in IVR and its impact on skill transfer and self-efficacy.

5.6 Conclusion

In conclusion, this preregistered study involving 130 participants investigated the effectiveness of adaptive immersive virtual reality (IVR) training compared to fixed IVR training in the context of the buzz wire task. Two primary hypotheses were examined, expecting adaptive training to lead to higher improvements in (1) performance and (2) self-efficacy compared to fixed training. The study was unique in its use of preregistration and large sample size, which are relatively uncommon in IVR research.

Despite the anticipation that adaptive training would yield better training outcomes, no significant differences in performance improvements were observed between adaptive and fixed training conditions, similar to findings by Zahabi and Abdul Razak (2020). One possible explanation for this outcome is that the fixed training condition was initially adjusted to participants' VR pretest performance, thus reducing the difference with adaptive training. Future research may consider altering the adaptation logic to provide more training in areas where participants feel more confident.

The study further demonstrated that IVR training had a significant effect on performance improvement both inside IVR (from VR pre- to VR post-test) and outside IVR in the physical transfer test. However, the training transfer effect was stronger within the IVR context than in the real world, which might be due to the contextual switch from IVR to the physical environment. Although improvements in both accuracy and speed were observed in IVR, only accuracy improvements were carried over to the real world, with speed performance worsening. Future research should explore whether these effects persist over extended periods of training or after longer intervals between tests.

In future studies, researchers could consider addressing limitations by incorporating longer training and retention tests to examine long-term effects and investigating transfer to different tasks using the same motor skills to better understand skill generalizability. Additionally, refining self-efficacy measurements to minimize biases in confidence assessment and adjusting the experimental design to balance focus on both mistakes and successes may lead to a better understanding of the effectiveness of adaptive training in IVR.

By employing a rigorous experimental design and a relatively large sample size, this study contributes valuable insights to the literature on adaptive training in IVR. The findings highlight the potential of IVR training to improve real-world performance, while also emphasizing the need for further research on optimizing adaptive training methods and understanding the factors that influence training transfer from virtual to physical environments.

Chapter 6 – Discussion & Conclusion

This chapter discusses the findings from the previous chapters and helps answer the original research questions of this dissertation. See Table 20 for a summary of the main results from the three experiments. For each research question, the primary findings are detailed, compared and contrasted with findings from literature and finally a few key take-aways are mentioned for inspiring further research. Subsequently, a sub-section reflecting on the methods developed during the dissertation and advice for researchers exploring the domain of IVR for industrial skills training is discussed, followed by the conclusion.

Table 20 Summary of experiments, study size and findings on performance improvements and arousal

Experiment No.	Experiment conditions	Average participants per condition	Total participant count	Performance improvements	Relation between arousal and performance
1	IVR vs physical training	44	87	No significant differences between IVR and physical training conditions	Participants with *lower* performance improvements exhibited higher arousal levels
2	Kinesthetic vs vibrotactile vs no haptic (visual only) feedback	24	73	No significant differences between Kinesthetic, vibrotactile and no haptic feedback conditions	Participants with *higher* performance improvements exhibited higher arousal levels
3	Adaptive vs fixed training	65	130	No significant differences between adaptive and fixed training conditions	-

6.1 What is the current state of the art in academic literature and industry practise regarding skills training using IVR?

The literature review, which examined 78 papers on immersive VR training, found that procedural (45%) and perceptual motor skills (33%) were the most common types of skills trained, with decision-making (17%) and spatial skills (5%) being less represented. A majority

(62%) of the papers followed a between-subjects experiment design. Performance metrics collected primarily focused on errors (accuracy) and task completion time (speed), while subjective metrics included cybersickness, usability, task load, immersion, and presence. Most of the papers did not use biosensors (82%) or haptic feedback (65%). However, over half (51%) compared IVR training's effectiveness with other modalities, typically finding it to be as effective or more. The majority of IVR applications (82%) were found to be suitable for remote training considering the COVID-19 pandemic.

Danish industry case studies revealed that training cost, scarce resources, and downtime were significant factors that made investing in IVR training reasonable for companies. Other motivating factors included safety, reusability, standardization, and mobility. In terms of design and technology, no specialized haptic devices or other apparatus were used to augment the training experience, although some reported intentions to apply peer cooperation. In DSB and Siemens-Gamesa use cases, haptic feedback was utilized, but only for non-motor skill-related information, such as collision with walls or task completion.

Recent literature reviews have largely supported the findings of our study. Abich et al. (2021) corroborated that IVR-based training is generally effective for enhancing performance in various skills. Meanwhile, Xie et al. (2021) and Renganayagalu et al. (2021) pinpointed the potential of haptic feedback and multi-user training, as well as identified the lack of pedagogical theories in the development of VR training studies. The findings from these studies are expected to remain relevant for the foreseeable future due to several factors. Firstly, the ongoing evolution of IVR technology along with technologies like haptics and biosensors, and their increasing accessibility are likely to drive further adoption and integration into industrial skills training. Furthermore, the potential benefits of IVR training, such as cost reduction, resource efficiency, and enhanced safety, will continue to attract industry attention and investment. As the understanding of pedagogical theories in IVR training grows, more robust and effective IVR training programs will emerge, further reinforcing the relevance of these findings in the years to come.

In summary, the current state of IVR training in both academic literature and industry practice showcases its efficacy and potential, while also emphasizing areas where further research and development are necessary.

6.2 Is IVR training effective compared to physical training?

The literature review found that in cases where IVR is compared to other training modalities (desktop VR, physical, video), it is effective in most of the cases. We also proceeded to test this by designing the experiment outlined in Chapter 3, to investigate the effectiveness of immersive virtual reality for training participants in the "buzz-wire" fine motor skill task compared to physical training. This between-subjects study was designed to assess and compare the performance improvements in both IVR and physical conditions (see Table 20 for a summary).

The findings revealed that both IVR and physical training resulted in statistically significant improvements in contact time from pre-training to post-training phases. To compare the effectiveness of the two training modalities, three metrics were defined: improvements in task completion time, improvements in contact time, and an Improvement Score, which combines the first two metrics. Statistical tests comparing these three metrics between IVR, and physical conditions demonstrated no statistically significant differences.

Existing literature supports these findings, with several studies showing IVR to be as effective as or more effective than physical training in various contexts. For instance, Hooper et al. found IVR to be more effective than physical training for hip arthroplasty surgery, Butt et al. observed the same advantage for catheter insertion training, although the advantage disappeared after a week, and Huber et al. reported IVR to be as effective as an 'augmented' VR condition. Our results also indicate that IVR training is as effective as physical training for the buzz-wire task, aligning with similar findings in other IVR skill training literature.

Implications for further research include exploring the long-term effects of IVR training compared to physical training, examining the transferability of skills between the two modalities, and investigating how different task complexities or skill domains might influence the effectiveness of IVR training. In conclusion, the current evidence supports the effectiveness of IVR training compared to physical training in various contexts.

6.3 What is the link between the physiological arousal level of the trainees and the effectiveness of IVR training?

In addressing this research question, the experiments in chapters 3 and 4 (see Table 20) aimed to investigate the relationship between physiological arousal and IVR training effectiveness for the buzz-wire task. The first experiment (detailed in chapter 3) compared IVR to physical training in a buzz-wire scenario. The experiment revealed that regardless of their training condition, participants with lower performance improvements exhibited higher arousal levels as compared to those with the highest improvements.

The second experiment as detailed in chapter 4 focused on the impact of different haptic feedback (kinesthetic vs. vibrotactile) on arousal levels and performance improvements for a buzz-wire IVR training scenario. Results showed higher arousal levels for participants receiving kinesthetic feedback compared to vibrotactile feedback. In addressing the research question, a weak correlation was found between performance improvements and arousal, indicating that as arousal increased during VR training, so did performance. This relationship was supported by further statistical trends, showing that participants with higher performance improvements were observed to have higher arousal that those with the least improvements.

Literature itself presents mixed findings on the relationship between arousal and training performance. Some studies suggest better performance to be linked to high arousal (Homer et al., 2019; Ünal et al., 2013), while others found it linked to low arousal (Kuan et al., 2018; Prabhu et al., 2010). It is possible that arousal and performance may not have a linear relationship but a more complex relationship which might involve other subjective psychological factors. For example, it was found that negative perception of increased arousal leads to anxiety (Ginty et al., 2022), and this has been shown to result in decreased performance in motor skill in domains like sports (Turner & Jones, 2018).

In summary, the first and second experiments investigated the relationship between performance and physiological arousal. These experiments reach different conclusions in terms of the relationship. i.e., the first experiment found participants with the highest improvements in performance to have lower arousal, while the second experiment found the highest improvement group to have higher arousal. But it does not necessarily imply a contradiction, as there are several factors which make the experiments different though they share similarities in terms of the core task (buzz-wire). The training itself was varied in terms of content and parameters like the workspace, hardware used, training difficulty levels among other factors (for more details refer to chapters 3 and 4). However, what is evident is that for the given training in both experiments, optimal arousal levels could be found. This is aligned to examples from literature that paint a more complex relationship between arousal and performance, instead of a simple linear relationship, and that there might be “optimal” arousal levels which are correlated with high performance (Arent & Landers, 2003; Wu et al., 2010). As certain arousal levels have been linked to better performance through these experiments, they hold implication for future research to investigate how changing the arousal levels for a given training task may lead to variations in performance improvements.

6.4 Can haptic feedback make IVR training more effective?

This research question is addressed with the help of the experiment described in chapter 4 which investigated the relationship between haptic feedback modality and IVR training effectiveness (Table 20). The experiment used a buzz-wire training scenario inside IVR where participants were given one of three different haptic feedback during training (i.e., when they made mistakes) – kinesthetic (directional force) feedback, vibrotactile feedback and no feedback. All the participants also received visual feedback when they made mistakes.

Findings from the experiment revealed that all the participants, regardless of the haptic condition (vibrotactile, kinesthetic or no feedback) showed performance improvements because of IVR training. However, it remained unclear which haptic modality was more effective in terms of improving training performance or if they were better than visual feedback alone. Additionally, findings from RQ3 indicated that kinesthetic feedback resulted in higher arousal compared to vibrotactile feedback. The findings also indicated that regardless of haptic condition, arousal among all the participants was correlated with increased performance. This suggests that future research could examine the potential for enhancing training performance by modulating arousal through haptic feedback. Care must be taken before generalizing the findings that kinesthetic or vibrotactile feedback is only as effective as no haptic feedback, since most motor tasks involve a mixture of visual sensing and haptic sensations, where in some cases haptics dominates (Adams et al., 2010), but it is generally found in literature that multimodal, i.e. combinations of visual, haptic and/or auditory feedback makes motor skill learning more effective (Sigrist et al., 2013).

In their literature review of vibrotactile feedback for motor skill training in VR, Islam and Lim (2022) found that having at least one modality of feedback was more effective than having no feedback at all. In our experiment, the “no feedback” condition (referred to as the “visual only” condition in chapter 4) had a visual component to it, and unlike all the cases in Islam et al’s review, the feedback was delivered in an immersive VR environment as opposed to a 2D display. Experiments may be devised to completely remove the visual aspect, however, in the case of the buzz-wire task this makes the training impractical and limits the applicability of the findings to motor skill training scenarios. The results of our experiment did not provide direct evidence on either vibrotactile or kinesthetic feedback proving to be more effective than no haptic feedback in performance measures or in presence measures. Therefore, for the buzz-wire training scenario used in the experiment, the haptic feedback methods are interchangeable, however, future research can explore if this finding holds for other IVR motor skill training scenarios. Literature shows that haptic feedback enhances the sense of presence and immersion in VR environments (Cooper et al., 2018; Gibbs et al., 2022), though it is possible the between subjects methodology used in the experiment did not give an accurate view of this. Future research can explore the relationship between haptics, presence and skill training as suggested by the CAMIL model which points to the benefits of increased presence on learning (Makransky & Petersen, 2021).

Recent examples of integration of haptic feedback devices in industrial training from SenseGlove[17] and HaptX[18] showcase some preliminary examinations in this regard, however, future research should verify their efficacy in improving training. Further research could explore the optimal combination of haptic modalities for industrial training scenarios and evaluate the effectiveness of adaptive haptic feedback based on individual arousal levels. Examining the long-term retention of skills acquired through different haptic feedback modalities could also provide valuable insights into the sustainability of IVR training.

6.5 What is the link between adaptive training and the effectiveness of IVR training?

The adaptive training experiment, detailed in chapter 5, explored the role of adaptation in improving the effectiveness of IVR training for the buzz-wire task. Participants in the experiment experienced either adaptive training or fixed training in IVR. In the adaptive condition, the training content was continuously adjusted to the self-efficacy and performance of the participants, whereas training content was adjusted once at the beginning of the training in the fixed condition. The results revealed no differences between the groups in terms of performance improvements on IVR and transfer tests. Across both conditions, participants showed performance improvements in speed and accuracy in the IVR test, while in the transfer test, performance in accuracy increased while speed decreased.

[17] https://www.senseglove.com/cases/volkswagen-commercial-vehicles/
[18] https://haptx.com/vr-training/

The literature on adaptive training using IVR and other media is minimal as noted by Zahabi and Abdul Razak (2020) and where available, they are pilot studies or concept papers. According to Zahabi and Abdul Razak (2020) and Kelley (1969), dynamically adapting the training content to fit trainee behaviour leads to better training outcomes. However, this hypothesis was not confirmed by our experiment which found participants who received fixed training to improve as much as those who received adaptive training. It should be noted that fixed training here does not imply that all the participants received the exact same training, but that the mix of speed and accuracy training sessions in the fixed condition was personalized to every participant at the beginning of training, according to their performance and self-efficacy measured after the first IVR test. It is highly conceivable that either of the two training conditions from the experiment would have been much more effective compared to completely non-adaptive training scenarios as seen in much of VR training literature and from our case study (see chapter 2). The methods used in this experiment: large sample sizes, the variation of IVR training to focus on either one of the component skills (speed and accuracy) and the use of self-efficacy measures for adaptation marks a point of departure from prior literature and can provide inspiration for the enhancement of IVR training systems in the future.

6.6 Reflection on methodologies and contributions

Over the course of the PhD study period, in service of answering the research questions, I have had and made use of opportunities to deepen my knowledge of the methodologies used to conduct research. Here, I discuss some reflections on both the methodologies used in the dissertation and the overall contributions.

- *Combination of diverse methodologies*: Skill training in VR has been explored considerably as seen from the literature review, similarly haptic feedback is a highly developed field and so is the literature on physiological arousal measurement through biosensors. However, there have been very few systematic explorations of the intersections of all these technologies, especially considering the benefit such an overlap would bring to improving IVR skill training. Furthermore, this exploration was preceded by providing a systematic understanding of existing academic literature and industry case studies, as there were lack of studies looking at the intersection of industrial skills training, haptic feedback, biosensors, and adaptive training. Hence, various methods and software had to be developed during the PhD study period to standardize the process of delivering IVR training, visual and haptic feedback, as well as the measurement of performance and biosignals. These methods are presented in more detail in sections 3.3, 4.3, 5.2. Additional methods were developed which will be published in the future, for example in the validation of the Polar Vantage V2[19] watch as a potential and more convenient replacement for the Polar H10 sensor. These methods simultaneously strengthened the experimentation while restricting its immediate application to a wider range of training scenarios. Future work should focus on extending this methodology to other industrial training scenarios, particularly for real world use cases.

- *Data collection and processing methods*: All the experiments used in the dissertation used iMotions as the base platform for storing both biosensor and performance data. The platform provided a validated and reliable tool supporting a wide range of biosensors, in addition to an open architecture which enabled for C# code in the Unity game engine to transmit training and test performance data to iMotions. However, due to limitations of the platform in analysing the type of customised data generated during this study, additional data processing pipelines had to be built using the ecosystem of data science packages in Python, primarily Jupyter notebooks, Pandas, Scipy, Seaborn among others. The Python ecosystem also provided access to tools for analysing heart rate variability and skin conductance biosignals. A key take-away and learning from course of this dissertation was the need for building reusable data processing pipelines that can manage large datasets such as those generated from biosensors, clean and pre-process it and finally generate insights into the data which can then potentially be used for scenarios like IVR training adaptation based on biosensor data.
- *Choice of skill training scenario*: The buzz-wire task was used in all three experiments, which involved participants holding in their right hand, a handle with a metallic loop attached to one end. The objective across all experiments was to pass the loop through a wire as fast as they can with the least mistakes (i.e., touching the wire). This choice of task was inspired by prior literature and served as a convenient platform as it was close in representation to many of the common fine motor skills required across industries like welding, surgery, or dentistry. Success in the task involved achieving both high speed and high accuracy, two metrics most found in IVR training literature.

[19] https://www.polar.com/en/vantage/v2/

Difficulty in both the test and training tasks were achieved by providing bends in the wire. Furthermore, this task also enabled the use of the left hand for attaching the electrodes for the Shimmer GSR+ sensor (for measuring skin conductance), as the palms of either hand were determined in a pilot study (and validated by literature) to be the best and the most convenient locations for measuring skin conductance. Thus, the buzz-wire task can be recommended as a promising testbed for conducting motor skill training experiments.

- *Use of biosensors*: Various sensors were considered in the measurement of physiological arousal. Electrodermal activity (skin conductance) was one established indicator of arousal, measured using the Shimmer GSR+, a lightweight wireless device previously validated in literature. Heart rate variability, another reliable indicator of arousal, was measured using the Polar H10, a medical grade ECG (electrocardiogram) device worn on the chest. Though literature backed up their use, pilot studies were performed to establish their efficacy in measuring arousal, particularly to check their reliability and functioning in conjunction with the iMotions platform. They were both found to be relatively convenient to use and analyse, in comparison to alternatives like EEG (electroencephalogram). However, along with opportunities, the use of these devices presented a few challenges. The Shimmer GSR+ required two electrodes to be connected to the index and middle fingers of the left hand and for the participant to hold the hand still to minimize signal artefacts caused by hand movement. This was quite challenging even though the experiment protocols had the participant moved the handle through the wire with their right hand. In the second experiment, which had participants restricted to a small workspace due to constraints of the haptic device, there were multiple instances of data loss. The Polar H10 on the other hand, did not have issues due to movement as it was quite robust as it was primarily designed for professional athletes. However, the protocols for wearing the H10 increased the experiment time considerably. Sensors using PPG (photoplethysmogram) can be considered for future IVR research since they are being integrated with HMDs like the Galea and the HP Reverb G2, however they need to be validated as the underlying technology generates heartrates (beat per minute) and not the more reliable indicator: RR intervals as measured by ECG devices. Future work should consider validating the use of more convenient measures of arousal, for example pupil dilation which is integrated into a few VR HMDs like the HTC Vive Eye Pro and Galea.

- *Use of haptic feedback in motor skill training*: In the first and third experiments where the only haptic sensation were provided to the participants when they made a mistake (i.e., virtual loop touched the virtual wire), additional visual feedback had to be provided so that they would not get confused. This is an issue affecting IVR training of motor skills and is also observed in the training of procedural skills. To overcome the lack of realistic haptic feedback in current generation IVR hardware, the third experiment employed the Geomagic Touch kinesthetic feedback device, which is essentially a robot, and a customized handle which delivered directional vibrotactile feedback (in contrast to the unidirectional feedback given by IVR controllers like the Oculus Touch). These types of specialized hardware present various challenges: development of software is restricted by the features available in off the shelf solutions. For example, customised haptic middleware had to be developed for the second experiment as the existing Unity plugins from Geomagic were insufficient for simulating the buzz-wire task. They also present challenges in the form of limited workspaces and the dangers of instabilities in the devices. Though the second experiment was inconclusive in measuring a difference between kinesthetic and vibrotactile feedback, the decision on using haptic devices for IVR skill training will ultimately be case specific, considering the complexity of the device, the complexity of development and measurable benefits of the device in enhancing training.
- *Generalizing findings from the experiments*: To increase internal validity, all three experiments facilitated distraction free, noise free training environments to minimize variability. The participants were randomly assigned to experimental conditions, and pre/post tests were conducted. This strengthened the internal validity of the experiment; however, it limits the external validity, i.e., will the findings from the experiments hold in real world scenarios? This presents some challenges as well as some opportunities, especially in relation to the investigation of the relationship between task performance and stress levels as represented by physiological arousal. For example, future research can extend controlled experiments as presented here to investigate the effect on stress levels and task performance, from simulating the noise and visual complexity of real-world scenarios (for example, a factory floor) in the experimental setup. Partial

replication is one method of establishing external validity through controlled experiments. For example, certain features of the IVR experiments used in this dissertation like the use of biosensors and haptics maybe replicated in another skill training scenario, and if the same findings are observed, external validity might be established.

- *Measuring effectiveness of IVR training*: All three experiments had a pretest and post-test which helped in measuring objective improvements in performance after training. But how does one establish that the improvement in performance is a result of the training and not just the effect of performing the test task twice? Literature does point to deliberate practice during training to be the biggest contributor to expert level performance (Ericsson & Charness, 1994), so it is likely that the improvements in performance seen across experiments is due to IVR training. Future studies might also explore this question by having an additional placebo group for every experiment that did nothing during training, so that practice effects are controlled. Another factor which needs to be considered in measuring IVR effectiveness is the span of the experiment itself. The current setup across all three experiments of having a test immediately before and after IVR training helped in measuring its short-term effects, however, further experiments should explore the long-term effects of IVR training by conducting retention tests after the span of a week or two.
- *Training strategies*: For all the experiments, there were no training instructions given, for example in optimizing the position of the handle or body position. The participants figured out optimal strategies in improving performance on their own. This was deliberate as the main research questions focused on the effect of media, haptic feedback and adaptive training, and to avoid potential interaction effects between media and instructional methods. Further research needs to be performed, particularly in the industrial context to investigate various instructional strategies.

Future work can build upon the findings of this dissertation. These may include expanding the methodology to other industrial training scenarios, validating more convenient measures of arousal, exploring the impact of different instructional strategies, and investigating the long-term effects of IVR training on skill retention. For eventual and wider acceptance of such techniques in the industry, it is essential that the complexity of such undertakings should be smoothened and standardized, especially for technologies like biosensors and haptics.

6.7 Conclusion

In conclusion, this dissertation has explored academic literature and the industry use cases around IVR based skill training, and used these to synthesize controlled experiments which investigated the factors affecting IVR training effectiveness: primarily physiological arousal, haptic feedback, adaptive training along with subjective measures like task load, self-efficacy, presence, and enjoyment. In addition, the dissertation provides guidelines and inspiration for the design of motor skill training experiments with the help of biosensors, IVR and haptic feedback. There is a need to explore the methodology used in this dissertation for more specific use cases from industry, while also being mindful of the potential drawbacks of being too context specific. Overall, the findings of this dissertation suggest that IVR has great potential as a tool for enhancing motor skill acquisition and transfer, and there is much to be gained from continued exploration and development in this area.

Appendix

List of videos

Experiment 1: https://tinyurl.com/buzz-wire-vr-1

Experiment 2: https://tinyurl.com/buzz-wire-vr-2

Experiment 3: https://tinyurl.com/buzz-wire-vr-3

Experiment 1 – Supplementary tables and figures

Table 21 – Physiological metrics, their source, and their relation to changes in arousal

Physiological Arousal Metric	Signal Source	Relation with arousal
Skin Conductance (SC)	EDA	SC↑ – Arousal ↑
Skin Conductance Response Amplitude (SCRAmp)	EDA	SCRAmp↑ – Arousal ↑
Skin Conductance Response Peaks Rate (SCRPeaks)	EDA	SCRPeaks↑ – Arousal ↑
Heart Rate (HR)	ECG	HR ↑ – Arousal ↑
Inter-Beat Interval (IBI)	ECG	IBI ↓ – Arousal ↑
Root Mean Square of Successive Difference (RMSSD)	ECG	RMSSD ↓ – Arousal ↑
Standard Deviation of NN Intervals (SDNN)	ECG	SDNN ↓ – Arousal ↑
Normalized High-Frequency Component (HFN)	ECG	HFN ↓ – Arousal ↑
LF/HF (Low Frequency/High Frequency) Ratio	ECG	LF/HF Ratio↑ – Arousal ↑

Table 22 Presence Questionnaire.

1. During the training session, I forgot that I was in a lab.
2. The training session totally filled my mind.
3. During the training session, I was very captivated by what was presented to me.
4. I felt like I really was present ("was there") during the training.
5. When the training session was over, I felt like I was back from a journey.
6. During the training session, I was not conscious of the room setup (assistant, speaker, chairs…).
7. During the training session, I lost the notion of time.
8. During the training session, I was living what I was seeing as if it was happening to me for real.
9. I lived the experience of performing the training intensely.
10. During the training session, I often thought of something else. (Inverted Question)
11. I felt more like a participant than spectator of the training.
12. I had to force myself to stay concentrated on the training session. (Inverted Question)
13. I always had in mind the fact that I was in a lab. (Inverted Question)
14. I was reacting to everything I was seeing as it was real.

Table 23 Immersion Questionnaire

The motor skill training experience …
1. Makes me feel immersed
2. Gives me the feeling that time passes quickly
3. Grabs all of my attention
4. Gives me a sense of being separated from the real world
5. Makes me lose myself in what I am doing
6. Makes my actions seem to come automatically
7. Causes me to stop noticing when I get tired
8. Causes me to forget about my everyday concerns
9. Makes me ignore everything around me

10.	Gets me fully emotionally involved
11.	Captivates me

Table 24. Performance metrics for the VR and Physical conditions: IS (Improvement Score), TCT-I (Improvement in Task Completion Time), and CT-I (Improvement in Contact Time)

Performance Metric	VR (Mean ± SD)	Physical (Mean ± SD)	p-value
IS	11.18 ± 5.11	10.79 ± 3.67	0.343
TCT-I	1.33s ± 8.57s	-0.83s ± 7.37s	0.286
CT-I	1.24s ± 2.04s	1.06s ± 1.14s	0.474

Table 25 Presence Scores

Presence Item	Physical (Mean ± SD)	IVR (Mean ± SD)	p-value
P1	3.19 ± 1.77	3.76 ± 1.63	0.0623
P2	4.88 ± 1.53	5.16 ± 1.22	0.2041
P3	5.31 ± 1.2	5.58 ± 1.27	0.1179
P4	5.43 ± 1.04	5.04 ± 1.45	0.2006
P5	3.45 ± 1.64	4.07 ± 1.54	**0.0359****
P6	3.12 ± 1.6	4.27 ± 1.9	**0.0018****
P7	4.62 ± 1.61	5.09 ± 1.5	0.0757
P8	4.93 ± 1.58	4.42 ± 1.86	0.1285
P9	4.79 ± 1.32	4.82 ± 1.4	0.4068
P10	3.86 ± 1.63	4.71 ± 1.18	**0.0042****
P11	5.74 ± 1.13	5.53 ± 1.53	0.4438
P12	4.07 ± 1.63	4.58 ± 1.44	0.0512
P13	2.81 ± 1.67	2.84 ± 1.58	0.4398
P14	5.43 ± 1.35	4.67 ± 1.73	**0.02348****
Combined	4.4 ± 0.79	4.61 ± 0.93	**0.0736***

Table 26. Immersion Scores

Immersion Item	Physical (Mean ± SD)	IVR (Mean ± SD)	p-value
I1	4.9 ± 1.39	5.02 ± 1.37	0.2899
I2	4.95 ± 1.34	5.36 ± 1.28	**0.0439****
I3	5.6 ± 1.19	5.64 ± 1.26	0.3436
I4	3.64 ± 1.64	4.78 ± 1.51	**0.0007****
I5	4.48 ± 1.78	4.64 ± 1.61	0.3528

I6	3.88 ± 1.71	4.33 ± 1.46	0.1225
I7	3.93 ± 1.79	4.53 ± 1.5	0.0501
I8	4.79 ± 1.69	4.96 ± 1.57	0.3224
I9	4.14 ± 1.57	4.87 ± 1.24	**0.0114****
I10	4.38 ± 1.65	4.71 ± 1.41	0.1748
I11	5.29 ± 1.2	5.51 ± 1.2	0.1461
Combined	4.54 ± 0.99	4.94 ± 0.98	**0.01751****

Table 27 NASA-TLX Scores

Item	Physical (Mean ± SD)	IVR (Mean ± SD)	p-value
Mental Demand	11.52 ± 4.55	12.1 ± 4.35	0.2738
Physical Demand	9.95 ± 5.9	10.22 ± 5.76	0.4008
Temporal Demand	11.71 ± 4.33	9.16 ± 4.83	**0.006****
Performance	11.38 ± 4.38	12.4 ± 4.6	0.1123
Effort	14.52 ± 4.84	14.49 ± 4.35	0.4073
Frustration	10.62 ± 5.66	10.82 ± 5.84	0.4257
NASA-TLX Score	11.62	11.53	0.379

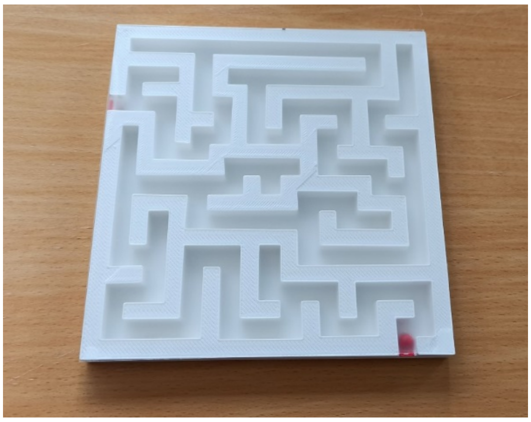


Figure 57. Distractor maze task

Buzz-wire pilot study paper

Investigating motor skill training and user arousal levels in VR : Pilot Study and Observations

Unnikrishnan Radhakrishnan* Department of Business Development and Technology Aarhus University

Alin Blindu† Aarhus University

Francesco Chinello‡ Department of Business Development and Technology Aarhus University

Konstantinos Koumaditis § Department of Business Development and Technology Aarhus University

Abstract

Virtual Reality (VR) for skill training is seeing increasing interest from academia and industry thanks to the highly immersive and realistic training opportunities they offer. Of the many factors affecting the effectiveness of training in VR, the arousal levels of the users merit a closer examination. Though subjective methods of measuring arousal exist in VR literature, there is potential in using cost-effective sensors to directly measure these from bio-signals generated by the nervous system. We introduce the design of preliminary observations from a pilot study exploring user's arousal levels and performance while executing a series of fine motor skill tasks (buzzwire tracing). Future directions of the work are also discussed.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual Reality;

1 Introduction

VR skill training systems offer several affordances such as providing familiarity with the associated real-world environment, increasing the salience of particular cues in the environment, allowing learners to rehearse specific sequences of actions with a high level of physical and psychological fidelity, and enabling immediate feedback and broad accessibility in a safe and controlled environment [3] [4]. Though VR based skill training is increasing in popularity across various domains, there is a need for a closer examination of the link between the user's arousal levels and training performance. Findings from such investigations could potentially be used to design adaptive VR training systems [2]. Compared to post-exposure subjective measures of arousal which rely on the user's memory of the task and comprehension of questions, objective measures of arousal are a function of the body's autonomic nervous system, which produces measurable responses reflecting the user's mental state. This includes changes in skin conductance (due to sweating) and heart rate parameters, among others which can be measured by various off-the-shelf sensors. Objective bio-signal data also allows for a more fine-grained look at variations in arousal across tasks with varying difficulty levels.

2 Experimental Setup

The buzzwire levels proposed were replicated both in VR and physical environments. The physical setup consisted of metal wires bent to fit the design of the different levels (see Fig. 1) while the VR setup is visualised in a VR environment using the Unity3D game engine and Oculus SDK. In the physical condition, the user is in a seated position with the dominant hand holding the Oculus controller enclosed in a 3D printed handle [5] coupled to the circular ring (of diameter 2.5 cm) as in Fig. 2(d). In the VR scenario, the user is in a seated position wearing an Oculus Quest and has one Oculus controller in the dominant hand.

*e-mail: unnik@btech.au.dk
†e-mail:alinblindu@gmail.com
‡e-mail:chinello@btech.au.dk
§e-mail:kkoumaditis@btech.au.dk

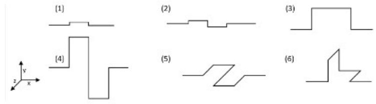


Figure 1: Buzzwire tracing task levels with (1) to (4) in the xy plane, (5) in the xz plane and (6) in all three axes.

The proposed buzzwire tracing task involves the user moving the circular ring across a thin wire (as in Fig. 2(a)), without touching it. The wire is bent at certain locations providing the challenge of maintaining proper hand-eye coordination while executing the task. This scenario has been used previously in VR and non-VR training scenarios [1] [6]. In order to explore the variation in arousal levels to the difficulty levels of the task, 6 difficulty levels were designed (see Fig. 1) to select a few representative levels for future experiments with larger populations. The total lengths of all 6 levels were kept at a constant 48cm.

The levels are designed as follows: level 1 has a bend with a height of 1.25 cm so that the user can pass the ring through x-axis (see Fig. 1); levels 2, 3 and 4 require hand rotations between the x and y axes. Level 5 changes the plane of motion (from xy to xz) while level 6 requires movement in all three axes.

An Arduino-based program detects when the ring touches the wire, and immediately switches on a red LED (Fig. 2(a)) to glow while simultaneously transmitting the data to the PC where a Unity3D script reads the data and triggers a short vibration (via the Oculus controller) and a sharp tone (via a Bluetooth-headphone) for each error. In the VR condition, Unity3D mesh-colliders are used to detect collisions between the ring and wire. For the VR setup, the audio and haptic feedback is the same as the physical condition, but for the visual feedback, a virtual light source in Unity3D is created which emits red light.

2.1 Bio-signal and Performance Measures

A Shimmer GSR+ (Galvanic Skin Resistance) unit (Fig. 2(c)) on the non-dominant hand (resting on the table) is used for measuring skin conductance (SC), and the Polar H10 chest strap is used for measuring the heart rate. Though SC can be recorded from any part of the skin, certain locations on the body are reported by literature to be better for recording higher quality data [7]. This was verified during a pretest comparing the quality of signals from different locations on the body, among which the fingers were identified after testing as being the most reliable. Ag/AgCl electrodes are placed on the index and middle fingers of the non-dominant hand, while the dominant hand is used to move the ring around the wire. The non-dominant hand is kept stationary throughout the experiment as GSR is affected by motion artifacts. Timestamped data from both Shimmer and Polar devices were collected by the iMotions software for later processing and analysis. Established measures of buzzwire

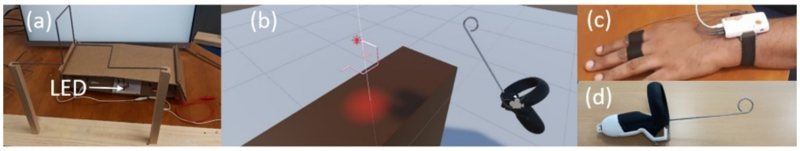


Figure 2: (a) Physical setup of level 6, (b) VR setup of level 6, (c) Shimmer GSR+ sensor, (d) Oculus controller with ring attached.

tracing (and similar fine motor skill tasks) like task completion time and count of errors are generated based on the timestamped data from iMotions during the data processing stage.

3 Preliminary Results and Analysis

Twelve participants volunteered to be part of the pilot study, 6 in the physical condition and 6 in the virtual reality condition, with three trials per level. The physical experimentation was performed with the biosensors, but the VR experiment was affected by Covid-19 restrictions and volunteers had access to a VR headset in an open area. The number of mistakes and task completion time were measured for both conditions.

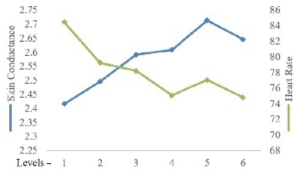


Figure 3: Average Skin Conductance (μS) & Average Heart Rate data across levels 1 to 6 (physical condition).

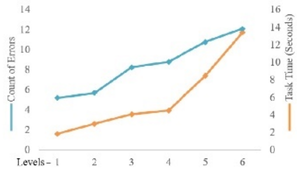


Figure 4: Average Count of Errors & Task Time (seconds) across levels 1 to 6 (physical condition)

Fig. 3 summarizes the results from the physical training tests. SC levels continue to climb on average for the 6 participants, though there is an unexpected drop from level 5 to level 6. HR drops on average from levels 1 to 6, though there is a slight climb from level 4 to 5 beating the downward trend. On the other hand, as shown in Fig. 4, task completion time and count of errors show a steady climb from levels 1 to 6, indicating that the initial reasoning for the design of the difficulty levels in the levels is sound. The climb observed in bio-sensor data for level 5 might be due to the novelty and challenge of tracing in a different plane compared to the previous levels. The data from the 6 participants in the VR condition (count of errors and task completion time) also confirmed these observations with a steady increase from levels 1 to 6.

Interviews and observations of both the physical and VR pilots suggest constructive changes and further research questions be explored. These include the role of decision making in the learning of fine motor skills since more than one participant mentioned the cognitive challenge of finding optimal solutions to levels 5 and 6. Though the buzzwires and the handle for the ring are similar in size across VR and physical conditions, there were differences in other aspects of the visual environment (table, background, etc). Some participants mentioned that they felt a slight delay in getting audio feedback when they commit a mistake, which is most possibly due to the Bluetooth connection. Two participants mentioned that they did not pay much attention to the visual feedback (LED) while for one person that was the primary modality.

4 Lessons learned and Future Work

The pilot study showed that (a) the bio-sensor setup (Shimmer GSR+ on the fingers and Polar H10 on the chest) generates reliable data, (b) the SC levels scaled up with the difficulty levels while the HR levels showed the opposite trend (c) the position of the LED (visual feedback) may have to be changed to a more prominent position as it was ignored by certain participants (d) the VR and physical experimental setups need to be designed to match each other closely as possible. Further investigation with a larger sample of participants needs to be done to conclude the relation between arousal levels as measured by the bio-sensors and training performance. The pilot study also sets the stage for robust experiments in the future to explore the role of audio/visual/haptic feedback (by comparing to the absence of it) and biofeedback (by allowing the user to see some visualization of their bio-sensor data) in affecting training performance across VR and physical conditions. More fine-grained analysis of bio-sensor data will also be explored since the pilot study only analyzed the skin conductance levels and not the skin conductance responses (peaks) that follow stimuli (in this case, errors and feedback) [7].

References

[1] C. Christou, D. Michael-Grigoriou, D. Sokratous, and M. Tsiakoulia. Buzzwirevr: An immersive game to supplement fine-motor movement therapy. 11 2018. doi: 10.2312/egve.20181327

[2] K. Koumaditis , F. Chinello, and S. Venckute. Design of a virtual reality and haptic setup linking arousals to training scenarios: A preliminary stage. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–2, 2018. doi: 10.1109/VR.2018.8446528

[3] K. Koumaditis, F. Chinello, P. Mitkidis, and S. Karg. Effectiveness of virtual versus physical training: The case of assembly tasks, trainer's verbal assistance, and task complexity. *IEEE Computer Graphics and Applications*, 40(5):41–56, 2020. doi: 10.1109/MCG.2020.3006330

[4] G. Makransky, S. Borre-Gude, and R. E. Mayer. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6):691–707, 2019.

[5] Octimusocti. Knuckles oculus quest and rift s grip, May 2019.

[6] J. C. A. Read, S. F. Begum, A. McDonald, and J. Trowbridge. The binocular advantage in visuomotor tasks involving tools. *i-Perception*, 4(2):101–110, 2013. PMID: 23755355. doi: 10.1068/i0565

[7] M. van Dooren, J. H. Janssen, et al. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior*, 106(2):298–304, 2012.

SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Adaptive study pilot paper

Adaptive Immersive VR Training Based on Performance and Self-Efficacy

Lasse F. Lui * Unnikrishnan Radhakrishnan Francesco Chinello Konstantinos Koumaditis

Extended Reality and Robotics Lab (xR^2)
Department of Business Development and Technology
Aarhus University

Abstract

Effective training requires that the training experience fits the ability and confidence (i.e., self-efficacy) of the trainee. Specifically, the individual's self-efficacy should ideally be slightly higher than the difficulty of a given task. A significant benefit of immersive virtual reality (IVR) is the potential to utilize measures of trainee behavior to continuously adapt the training content to the individual. However, a major challenge involves the identification of relevant measures that can be used to adapt training content in a way that increases training output. The current paper aims to inspire further research on adaptive IVR training by describing the design and development of a study on adaptive IVR training where the use of self-efficacy measures for adaptation marks a point of departure from prior literature. The design of the proposed study is informed by analyses of results from previous IVR studies.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality;

1 Introduction

Virtual Reality (VR) refers to technologies that enable the experience of simulated reality in the form of computer-generated imagery, sounds, tactile sensations, etc. VR technologies range from desktop PC-based VR to CAVE displays and Head Mounted Displays (HMDs) [25]. The latter technologies are referred to as Immersive VR (IVR) as they immerse the user in the virtual environment by enabling wide fields of view, head tracking, and varying levels of interaction between the user and the environment. Skill training in IVR is becoming popular in the industry spanning multiple domains [20]. Interest in IVR-based skill training is primarily motivated by its potential to, among other factors, improve the learning curve, reduce training cost and time, and increase portability with safety [1]. Although previous re-views highlight most studies show positive effects of IVR training [20], the mixed and often conflicting nature of results indicates further explorations are needed to investigate when IVR training does and does not work [6]. It should be noted that non-performance related measures are also considered when measuring the effectiveness of IVR training. E.g., physiological arousal during IVR training has been linked to improvements in performance [19,21] and measures such as self-efficacy and presence (the subjective experience of being in an immersive VR environment) lead to better training outcomes, particularly when the affordances of IVR align with the used instructional method [15].

In their systematic review of IVR-based training, Abich et al. [1] identified the kinds of knowledge, skills, and abilities improved by IVR training with three distinct categories identified: psychomotor performance, knowledge acquisition, and spatial ability. Specifically, psychomotor performance in tasks such as assembly and surgery benefitted from IVR training. In their review, Radhakrishnan et al. [20] found fine and gross motor skills to dominate publications from the medical domain. The effectiveness of IVR training is determined by varying aspects of the training environment including, among other factors, level of immersion, training method, haptic feedback, and gamification [20].

*e-mail:lui@btech.au.dk

Both Abich et al. [1] and Radhakrishnan et al. [20] found time and accuracy (based on mistakes committed) as common performance metrics used to evaluate the effectiveness of VR training. Though most examples from the literature are domain-specific, especially those from the medical domain, non-domain specific examples also exist, such as assembling 3D objects, which employ a mix of procedural and fine motor skills [5, 10, 18]. The current paper extends previous non-domain specific use cases by utilizing a buzzwire fine motor skill training scenario [7, 19]. In these experiments, participants are asked to move a loop across a wire as fast as they can while making as few mistakes as possible (i.e., avoid the loop touching the wire; see Fig. 1).

Christou et al. [7] designed an IVR buzzwire task targeting movement rehabilitation among stroke-affected patients. They used this in two experiments, where the first one (among able-bodied participants) showed that hand dominance (left vs right) could be distinguished based on task performance data (such as speed and mistakes). In the second experiment, conducted among stroke pa-tients, they were able to distinguish between the weak and the strong arms of each participant. They finally proposed an IVR buzzwire training system comprised of different levels of increasing difficulty.

Radhakrishnan et al. [19] investigated performance improvements in a physical buzzwire task among participants randomly placed into either physical training or IVR training condition (see Fig. 1). Results showed that both conditions improved task performance. Participants in the IVR condition reported more enjoyment and higher temporal demand (measured using NASA-TLX) compared to those in the physical training condition. Physiological arousal was measured using both skin conductance and heart rate sensors, and participants in the IVR condition were less aroused than those in the physical training condition. Furthermore, participants with the lowest arousal levels had higher improvements in performance than those with the highest arousal levels across conditions. Lastly, IVR training led to a statistically significant increase in self-efficacy from pre-training to post-training ($p = .016$) whereas physical training did not ($p = .545$). Thus, although IVR training did not lead to higher performance than physical training, it did lead to a significantly higher increase in self-efficacy suggesting IVR-based training is especially fruitful in supporting the trainee's self-efficacy. Based on these results, the current study explores the role and potential of self-efficacy in fine motor skills training using the buzzwire task.

Self-efficacy (SE) refers to one's perceived capabilities for learning or performing actions [22]. According to the Social Cognitive Theory formulated by Bandura [2], SE is among the most important factors in learning and training. Ideally, the individual's level of SE should be slightly higher than the difficulty of the task at hand, meaning they have the feeling of being able to complete the task without finding it too easy or too difficult for their current ability. A major source of SE is based on performance accomplishment [24], i.e. when the individual experiences success in performing an action

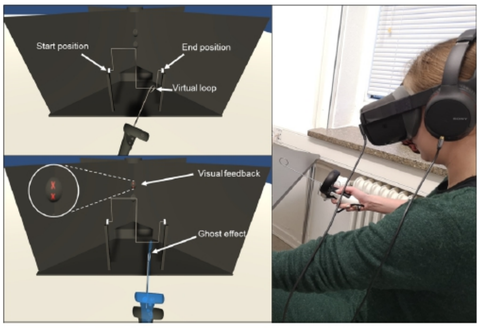


Figure 1: Right: Study participant practicing buzzwire task in VR. Top left: Virtual loop being moved across a virtual buzzwire, Bottom Left: Ghost loop appears when contact is made, and the 'real' loop goes outside the wire. It disappears when the loop is placed back inside the wire. Figure is taken with permission from Radhakrishnan et al. [19].

or task, their level of SE will increase. IVR training could therefore benefit from aiming to strengthen not only performance but also the individual's confidence in performing the task at a satisfactory level (i.e., self-efficacy).

Previous research supports the importance of SE in learning and training (for more information, see the meta-analysis of 430 independent samples by Sitzmann and Ely [23]. However, few studies investigated the role of SE in IVR-based learning and training [12], with some support for the capacity of IVR to increase SE in laboratory safety training [14], crane training [24], and be as good as traditional Aviation safety training [4]. That said, studies rarely focus on how specific features of IVR can be used and changed to increase the individual's level of SE. A central implication of viewing training through the lens of SE is that every trainee is different and has a different level of SE for the task at hand [9]. Therefore, virtual environments that adapt based on the behavior of the trainee with the aim of increasing performance through heightened SE have the potential to create a better fit between individual levels of SE and aspects of the training content.

Most current studies do not use adaptation. In other words, the same training is provided to all trainees regardless of their cognitive load, physiological arousal, and expertise level [26]. Of the few studies on adaptive IVR, most implement simple forms of adaptability, often based on pre-evaluations before or at the beginning of the IVR experience [13, 26]. Although these are partly tailored to the individual, they do not include features that continuously monitor and adapt the learning content throughout the experience. In 1969, Kelley (1969) described adaptive training as requiring that *"[...] performance be continuously or repetitively measured in some way, and that the measurement be employed to make appropriate changes in the stimulus, problem, or task."* Based on a literature review, Zahabi and Abdul Razak [26] propose a framework for adaptive IVR training that incorporates performance measures and adaptive variables (i.e., changes in the training) based on adaptive logic (which changes the training based on performance measures). Inspired by Gerbaud et al. [8], Zahabi and Abdul Razak [26] describe two types of adaption based on the timing of the adaptive logic: parametering adaptation refers to adaptation occurring before the start of training, whereas dynamic adaptation refers to adaptation occurring during training. Based on such prior literature, the current paper utilizes adaptive immersive virtual reality (adaptive IVR) to refer to IVR-based applications with continuously adaptive con-tent based on the

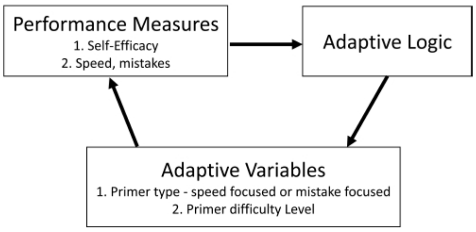


Figure 2: Design of the proposed adaptive IVR training system

real-time behavior of the user.

Adaptive IVR studies matching this definition are currently limited. For example, Kritikos et al. [11] used adaptive IVR to treat patients with arachnophobia by continuously adjusting the intensity of anxiety induction based on Electrodermal Activity (EDA) measures, whereas Muñoz et al. [17] used Electrocardiogram (ECG) and electro-encephalogram (EEG) measures to support calm-ness and cognitive readiness for military personnel training. However, the potential of adaptive IVR for fine motor skills training has not been investigated, which is the main purpose of the current paper. Nevertheless, Mariani et al. [16] provides inspiration in the form of adaptive training for robotic surgery skills in a non-immersive VR setup. In the study, participants received training in specific 'elementary' skills based on their performance in complex tasks during training. Despite the importance of the subjective evaluation of one's own competence, there are currently no examples from IVR literature which forms adaptive logic based on SE. The current paper focuses on the decisions related to design, development, and testing of an adaptive IVR application that continuously adapts training content based on dynamic measures of both SE and task performance of the trainee as shown in Fig. 2. The proposed adaptive system is further detailed in the Methods section. Overall, the aim can be summed up by the following three research questions:

RQ 1*: Does adaptive IVR training support performance in terms of speed and precision?*

RQ 2*: Does adaptive IVR training support the individual's self-efficacy?*

RQ 3*: Does self-efficacy play a role in the relationship between IVR training and performance?*

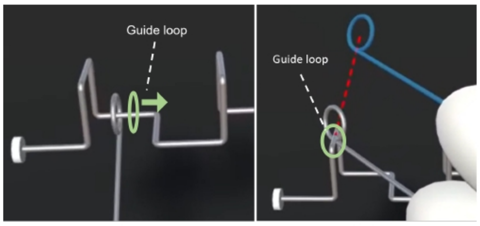
2 Methods

This section describes the considerations related to designing and preparing a lab experiment based on data from previous buzzwire experiments [19, 21].

2.1 Experimental design

Approximately 120 participants are planned to be recruited for a between-subjects experiment. The study will be preregistered using AsPredicted [1] and preregistration, study materials, and data will be made available on the Open Science Framework (OSF) [2]. As with previous buzzwire experiments, participants will complete a pre-test and post-test on a physical version of the buzzwire task, in order to measure changes in performance. Self-reported SE data will also be collected before and after training. Performance is measured in terms of *task completion time* (TCT), the time it takes for the loop

[1] https://aspredicted.org/
[2] https://osf.io/



(a) (b)

Figure 3: (a) Speed focused priming. A green guide loop moves at a constant speed across the wire. (b) Mistake-focused priming. A green guide loop stays on the wire while a mistake is being made.

to be moved from one end of the wire to the other, and *contact time* (CT), the total time the loop is in contact with the wire (as a measure of errors). During the main part of the experiment, participants will complete 10 repetitions of an IVR buzzwire task. Repetitions are important in motor skill training as according to Bernstein et al. [3]: *"Repetitions of a movement or action are necessary to solve the motor problem many times and to find the best way of solving it."*. After each repetition, two items will be utilized to measure SE which informs the adaptive logic. Results from previous buzzwire studies (including 3 and 6 repetitions) led to the decision to use more repetitions to increase sensitivity.

It was decided to include two adaptive variables, priming type and priming difficulty, based on measures of SE and performance. In the current paper, priming refers to the guidance provided by the IVR training with the help of visual cues (elaborated in Sec. 2.2 and Sec. 2.3). To investigate the effect of adaptive training, a comparison will be made be-tween personalized IVR training (the control group), where the training content is based on trainee behavior from the pre-test, and adaptive training (the experimental group), where training content will continuously change based on changes the trainee's behavior.

Therefore, the type of primer and difficulty will exclusively be based on measurements from the pre-test for the control group (parametering adapta-tion), whereas performance and SE will be measured continuously to dynamically adapt the primer and difficulty throughout the training for the experimental group (dynamic adaptation). That way both groups receive a degree of adaptation, but only the experimental group will receive dynamic adaptations. The following subsections describe how the exploratory analysis was utilized to decide which measures and adaptive variables would make the best fit for the buzzwire IVR training.

2.2 Adaptive variable 1: Primer type based on self-efficacy

As described by Zahabi and Abdul Razak [26], adaptive systems should include adaptive logic based on relevant measures. Multiple regression analyses of the previous buzzwire studies indicated there may be individual differences in performance improvement. Specifically, TCT during the pre-test predicted improvements in TCT (i.e., the change from pre-test to post-test; $p < .001$, $B = 0.050$) but not CT ($p = .22$, $B = 0.018$), whereas the opposite was the case for pre-test CT, which predicted improvements in CT ($p < .001$, $B = 0.446$) but not TCT ($p = .50$, $B = 0.089$). This could indicate individual differences in the ability to perform the task either quickly or precisely predicts which type of improvement the individual trainee will achieve. Therefore, the present study will implement one of two types of priming as an adaptive variable during each IVR buzzwire repetition to either support speed- or mistake-focused training.

For mistake-focused priming, a guide loop is added during the task. The guide loop (Fig. 3b) is always located in the optimal position of the loop, i.e., as far from the wire as possible. This is to help the participant focus on making fewer mistakes by making it apparent when they are close to making a mistake (i.e., getting further away from the guide loop).

For speed-focused priming, a guide loop (Fig. 3a) is moving through the buzzwire on its own. This is to make the participant focus on performing the task as fast as possible by competing against the guide loop.

Since indications of two different types of trainees were identified, SE was chosen as a measure for the adaptive logic since, 1) prior research supports the relevance of matching training content based on the individual's SE, and 2) results of the study by Radhakrishnan et al. [19] suggested support for increasing SE through IVR training. Specifically, the participant will respond to two questions after each repetition of the buzzwire task (I feel confident that I can complete the task faster; I feel confident that I can complete the task with fewer mistakes). In the control group, most primers (approximately 80%) will match the type of performance the participant reports the least confidence in during the pre-test, whereas participants in the experimental group will receive the type of primer they reported the least confidence in during their last repetition of the task.

2.3 Adaptive variable 2: Primer difficulty based on performance

To support differences in baseline motor skills, priming difficulty is varied by changing the size of the loop in the mistake-focused priming and the speed of the guide loop in the speed-focused priming. Two performance measures were considered for the adaptive logic of difficulty: arousal and performance. Kritikos et al. [11] and Muñoz et al. [17] both used arousal (ECG, EDA, and EEG) in adaptive IVR. However, these studies both included stronger stimuli related to arachnophobia and military training compared to the training content in the buzzwire task. This was confirmed during exploratory analysis, where no clear pattern could be identified for using arousal to determine the adaptive logic in the buzzwire task. Instead, the performance in the pre-test was deemed more appropriate for the present design since there was a clear pattern of differences in individual performance (see multiple regression analyses described in Sec. 2.2) with a connection to the outcome measure.

In previous buzzwire studies, participant performance increased by an average of 2.25% (speed) and 3.3% (mistakes) for each repetition of the task. Based on this information, a constant difficulty will be set to match the pre-test performance for the control group followed by a constant increase in difficulty of 2% (speed) and 3% (mistakes) for each repetition. In contrast, the pre-test performance only affects the difficulty of the first repetition for the experimental group, where-after the difficulty will adapt depending on continuous measures of participant performance.

Overall, it was decided to include two measures for two types of adaptive variables to support different types of training during each buzzwire repetition: a speed- or mistake-focused primer, based on SE, and a personalized level of difficulty, based on performance.

2.4 Analysis

The main analyses of the study will be based on the three research questions:

Does adaptive IVR training support performance in terms of speed and precision?

Differences in performance will be measured as the change in performance from the pre-test to the post-test (i.e., the effect of IVR training). Previous studies utilized a relatively complex scoring system to account for participants performing either far above or far below the mean during the pre-test. Exploratory analysis suggested that changes in performance based on percentage improvements would be a better fit for this study.

Therefore, to account for differences in baseline fine motor skill, change in performance is measured in terms of percentage improvement. For example, completion times of 60 seconds (pre) and 54 sec-onds (post) is interpreted as the same 10% improvement as completions times of 30 (pre) and 27 seconds (post). Analyses will then compare the control and experimental group in terms of mean improvements in 1) time, 2) mistakes, and 3) combined improvement.

Does adaptive IVR training support the individual's self-efficacy?

The groups will be compared in terms of mean change in SE from before to after training. Whereas two items measuring SE after each repetition of the buzzwire task are used for the adaptive logic, statis-tical analysis of the change in SE will be based on the SE measure collected before the pre- and post-test. Since the SE measure consists of multiple items, each participant´s SE score is based on their mean rating of these items and change in SE will be calculated by subtracting pre-SE score from post-SE score. Secondary analysis will explore whether dynamic changes in SE during IVR training are different in the two groups.

Does self-efficacy play a role in the relationship between IVR training and performance?

The relationship between SE and training will be analyzed with multiple regression analysis predicting change in performance-based on experimental condition, SE, and their interaction. Further analysis will explore the potential mediating role of SE in the relationship between IVR-based training and improvement in performance. All main and secondary analyses will be preregistered at AsPredicted.org before data collection.

3 Conclusion

Adaptation presents opportunities to improve the efficacy of IVR training. While personalization using parameterized adaptive methods is predominant in the design of adaptive IVR literature, this paper presented the potential for dynamic adaptation to increase the effectiveness of training. While the current IVR adaptive training literature uses measures of task performance, cognitive load, and arousal to adapt the training, the use of self-efficacy measures for adaptive IVR training is lacking. Furthermore, there is evidence that adapting training difficulty to suit a trainee's self-efficacy leads to greater improvements in training outcomes. Three research questions were generated in order to investigate the effectiveness of an IVR fine motor skill training system that dynamically adapts training difficulty based on self-efficacy and trainee performance during training. Lastly, the planning and design of a controlled experiment were presented with the aim of answering these research questions by comparing training outcomes between IVR-based training with and without dynamically adapting content.

References

[1] J. Abich IV, J. Parker, J. S. Murphy, and M. Eudy. A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality*, 25(4):919–933, 2021.

[2] A. Bandura. Social cognitive theory: An agentic perspective. *Asian journal of social psychology*, 2(1):21–41, 1999.

[3] N. A. Bernstein, M. L. Latash, and M. T. Turvey. *Dexterity and its development*. Psychology Press, 2014.

[4] F. Buttussi and L. Chittaro. Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics*, 24(2):1063–1076, 2017.

[5] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse. Virtual training: Learning transfer of assembly tasks. *IEEE transactions on visualization and computer graphics*, 21(6):770–782, 2015.

[6] D. Checa and A. Bustillo. A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 79(9):5501–5527, 2020.

[7] C. G. Christou, D. Michael-Grigoriou, D. Sokratous, and M. Tsiakoulia. Buzzwirevr: An immersive game to supplement fine-motor movement therapy. In *ICAT-EGVE*, pp. 149–156, 2018.

[8] S. Gerbaud, V. Gouranton, and B. Arnaldi. Adaptation in collaborative virtual environments for training. In *International conference on technologies for E-learning and digital entertainment*, pp. 316–327. Springer, 2009.

[9] R. M. Klassen and E. L. Usher. Self-efficacy in educational settings: Recent research and emerging directions. *The decade ahead: Theoretical perspectives on motivation and achievement*, 2010.

[10] K. Koumaditis, F. Chinello, P. Mitkidis, and S. Karg. Effectiveness of virtual versus physical training: The case of assembly tasks, trainer's verbal assistance, and task complexity. *IEEE Computer Graphics and Applications*, 40(5):41–56, 2020.

[11] J. Kritikos, G. Alevizopoulos, and D. Koutsouris. Personalized virtual reality human-computer interaction for psychiatric and neurological illnesses: a dynamically adaptive virtual reality environment that changes according to real-time feedback from electrophysiological signal responses. *Frontiers in Human Neuroscience*, 15:596980, 2021.

[12] A. Lehikko. Measuring self-efficacy in immersive virtual learning environments: A systematic literature review. *Journal of Interactive Learning Research*, 32(2):125–146, 2021.

[13] W. Li, H. Huang, T. Solomon, B. Esmaeili, and L.-F. Yu. Synthesizing personalized construction safety training scenarios for vr training. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):1993–2002, 2022.

[14] G. Makransky, S. Borre-Gude, and R. E. Mayer. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6):691–707, 2019.

[15] G. Makransky and G. B. Petersen. The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3):937–958, 2021.

[16] A. Mariani, E. Pellegrini, and E. De Momi. Skill-oriented and performance-driven adaptive curricula for training in robot-assisted surgery using simulators: A feasibility study. *IEEE Transactions on Biomedical Engineering*, 68(2):685–694, 2020.

[17] J. E. Muñoz, A. T. Pope, and L. E. Velez. Integrating biocybernetic adaptation in virtual reality training concentration and calmness in target shooting. In *Physiological Computing Systems: International Conferences, PhyCS 2016, Lisbon, Portugal, July 27–28, 2016, PhyCS 2017, Madrid, Spain, July 27–28, 2017, PhyCS 2018, Seville, Spain, September 19–21, 2018, Revised and Extended Selected Papers 3*, pp. 218–237. Springer, 2019.

[18] M. Murcia-Lopez and A. Steed. A comparison of virtual and physical training transfer of bimanual assembly tasks. *IEEE transactions on visualization and computer graphics*, 24(4):1574–1583, 2018.

[19] U. Radhakrishnan, F. Chinello, and K. Koumaditis. Investigating the effectiveness of immersive vr skill training and its link to physiological arousal. *Virtual Reality*, pp. 1–25, 2022.

[20] U. Radhakrishnan, K. Koumaditis, and F. Chinello. A systematic review of immersive virtual reality for industrial skills training. *Behaviour & Information Technology*, 40(12):1310–1339, 2021.

[21] U. Radhakrishnan, L. Kuang, K. Koumaditis, F. Chinello, and C. Pacchierotti. Haptic feedback, performance, and arousal: A comparison study in a virtual reality motor skill training task. January 2023.

[22] D. H. Schunk and M. K. DiBenedetto. Self-efficacy theory in education. In *Handbook of motivation at school*, pp. 34–54. Routledge, 2016.

[23] T. Sitzmann and K. Ely. A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go. *Psychological bulletin*, 137(3):421, 2011.

[24] H. Song, T. Kim, J. Kim, D. Ahn, and Y. Kang. Effectiveness of vr crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction*, 122:103506, 2021.

[25] K. Tcha-Tokey, E. Loup-Escande, O. Christmann, and S. Richir. Effects on user experience in an edutainment virtual environment: comparison between cave and hmd. In *Proceedings of the European conference on cognitive ergonomics 2017*, pp. 1–8, 2017.

[26] M. Zahabi and A. M. Abdul Razak. Adaptive virtual reality-based training: a systematic literature review and framework. *Virtual Reality*, 24(4):725–752, 2020.

References

Abbas, Z.-A., & North, J. S. (2018). Good-vs. poor-trial feedback in motor learning: The role of self-efficacy and intrinsic motivation across levels of task difficulty. *Learning and instruction, 55*, 105-112.

Abbott, J. J., Marayong, P., & Okamura, A. M. (2007). Haptic virtual fixtures for robot-assisted manipulation. Robotics Research: Results of the 12th International Symposium ISRR,

Abich, J., Parker, J., Murphy, J. S., & Eudy, M. (2021, 2021/12/01). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality, 25*(4), 919-933. https://doi.org/10.1007/s10055-020-00498-8

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2010). Efficient visual recalibration from either visual or haptic feedback: the importance of being wrong. *Journal of Neuroscience, 30*(44), 14745-14749.

Adhanom, I. B., Al-Zayer, M., Macneilage, P., & Folmer, E. (2021). Field-of-View Restriction to Reduce VR Sickness Does Not Impede Spatial Learning in Women. *ACM Trans. Appl. Percept., 18*(2), Article 5. https://doi.org/10.1145/3448304

Aggarwal, R., Moorthy, K., & Darzi, A. (2004). Laparoscopic skills training and assessment. *British Journal of Surgery, 91*(12), 1549-1558. https://doi.org/10.1002/bjs.4816

Aggravi, M., Pausé, F., Giordano, P. R., & Pacchierotti, C. (2018). Design and evaluation of a wearable haptic device for skin stretch, pressure, and vibrotactile stimuli. *IEEE Robotics and Automation Letters, 3*(3), 2166-2173.

Akanmu, A. A., Olayiwola, J., Ogunseiju, O., & McFeeters, D. (2020, 09/2020). Cyber-physical postural training system for construction workers. *Automation in Construction, 117*, 103272. https://doi.org/10.1016/j.autcon.2020.103272

Alaker, M., Wynn, G. R., & Arulampalam, T. (2016). Virtual reality training in laparoscopic surgery: a systematic review & meta-analysis. *International Journal of Surgery, 29*, 85-94.

Allcoat, D., & von Mühlenen, A. (2018, 2018-11-27). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology, 26*(0). https://doi.org/10.25304/rlt.v26.2140

Almousa, O., Prates, J., Yeslam, N., Mac Gregor, D., Zhang, J., Phan, V., Nielsen, M., Smith, R., & Qayumi, K. (2019, February 1, 2019). Virtual Reality Simulation Technology for Cardiopulmonary Resuscitation Training: An Innovative Hybrid System With Haptic Feedback. *Simulation & Gaming, 50*(1), 6-22. https://doi.org/10.1177/1046878118820905

Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. 2016 IEEE aerospace conference,

Arent, S. M., & Landers, D. M. (2003). Arousal, anxiety, and performance: A reexamination of the inverted-U hypothesis. *Research Quarterly for Exercise and Sport, 74*(4), 436-444.

Ashiri, M., Lithgow, B., Suleiman, A., Blakley, B., Mansouri, B., & Moussavi, Z. (2020, 2020/04/01). Differences Between Physical vs. Virtual Evoked Vestibular Responses. *Annals of Biomedical Engineering, 48*(4), 1241-1255. https://doi.org/10.1007/s10439-019-02446-3

Ayala García, A., Galván Bobadilla, I., Arroyo Figueroa, G., Pérez Ramírez, M., & Muñoz Román, J. (2016). Virtual reality training system for maintenance and operation of high-voltage overhead power lines. *Virtual Reality, 20*, 27-40.

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators & Virtual Environments, 6*(4), 355-385.

Baceviciute, S., Mottelson, A., Terkildsen, T., & Makransky, G. (2020, 2020-04-21). Investigating Representation of Text and Audio in Educational VR using Learning Outcomes and EEG. CHI '20: CHI Conference on Human Factors in Computing Systems,

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. WW Norton & Company.

Bálint, L. A. P., Macías, L. H. P., & Althoefer, K. (2019). Virtual Reality Simulator for Medical Auscultation Training. In K. Althoefer, J. Konstantinova, & K. Zhang (Eds.), *Towards Autonomous Robotic Systems* (Vol. 11649, pp. 386-398). Springer International Publishing.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review, 84*(2), 191.

Bandura, A. (1986). The Explanatory and Predictive Scope of Self-Efficacy Theory. *Journal of Social and Clinical Psychology, 4*(3), 359-373. https://doi.org/10.1521/jscp.1986.4.3.359

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*(1), 307-337.

Bargh, J. A., & Cohen, J. L. (1978). Mediating factors in the arousal-performance relationship. *Motivation and Emotion, 2*, 243-257.

Barkokebas, R., Ritter, C., Sirbu, V., Li, X., & Al-Hussein, M. (2019, 2019-05-24). Application of Virtual Reality in Task Training in the Construction Manufacturing Industry. 36th International Symposium on Automation and Robotics in Construction,

Basalp, E., Wolf, P., & Marchal-Crespo, L. (2021). Haptic training: which types facilitate (re) learning of which motor task and for whom? answers by a review. *IEEE transactions on haptics, 14*(4), 722-739.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report, 13*(4), 544-559.

Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one, 11*(2), e0148037.

Biggs, S. J., & Srinivasan, M. A. (2002). Haptic interfaces. *Handbook of virtual environments*, 93-116.

Billings, D. (2012). Efficacy of adaptive feedback strategies in simulation-based training. *Military Psychology, 24*(2), 114-133.

Bin, F., Xi, Z., Yi, C., & Ping, W. G. (2019, 2019-08-09). Construction safety education system based on virtual reality. *IOP Conference Series: Materials Science and Engineering, 563*, 042011. https://doi.org/10.1088/1757-899X/563/4/042011

Binsch, O., Bottenheft, C., Landman, A., Roijendijk, L., & Vermetten, E. H. G. J. M. (2021, 2021/05/04). Testing the applicability of a virtual reality simulation platform for stress training of first responders. *Military Psychology, 33*(3), 182-196. https://doi.org/10.1080/08995605.2021.1897494

Bjork, R. A., & Whitten, W. B. (1974, 1974/04/01/). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*(2), 173-189. https://doi.org/https://doi.org/10.1016/0010-0285(74)90009-7

Blumstein, G., Zukotynski, B., Cevallos, N., Ishmael, C., Zoller, S., Burke, Z., Clarkson, S., Park, H., Bernthal, N., & SooHoo, N. F. (2020, 2/2020). Randomized Trial of a Virtual Reality Tool to Teach Surgical Technique for Tibial Shaft Fracture Intramedullary Nailing. *Journal of Surgical Education*, S1931720420300027. https://doi.org/10.1016/j.jsurg.2020.01.002

Botvinick, M., & Cohen, J. (1998, 2/1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756-756. https://doi.org/10.1038/35784

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bowman, D. A., & McMahan, R. P. (2007). Virtual Reality: How Much Immersion Is Enough? *Computer, 40*(7), 36-43. https://doi.org/10.1109/MC.2007.257

Bradley, M. M., & Lang, P. J. (1994, 1994/03/01/). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59. https://doi.org/https://doi.org/10.1016/0005-7916(94)90063-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology, 49*(1), 1017-1034.

Brooke, J. (1996, 1996). SUS: a "quick and dirty" usability scale. *Usability evaluation in industry*, 189.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning [Article]. *Educational Psychologist, 38*(1), 53-61. https://doi.org/10.1207/S15326985EP3801_7

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of cognition, 2*(1).

Budini, F., Lowery, M. M., Hutchinson, M., Bradley, D., Conroy, L., & De Vito, G. (2014). Dexterity training improves manual precision in patients affected by essential tremor. *Archives of physical medicine and rehabilitation, 95*(4), 705-710.

Buń, P., Górski, F., & Lisek, N. (2019). Comparison of Manual Assembly Training Possibilities in Various Virtual Reality Systems. In J. Machado, F. Soares, & G. Veiga (Eds.), *Innovation, Engineering and Entrepreneurship* (Vol. 505, pp. 398-404). Springer International Publishing.

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.

Butt, A. L., Kardong-Edgren, S., & Ellertson, A. (2018, 03/2018). Using Game-Based Virtual Reality with Haptics for Skill Acquisition. *Clinical Simulation in Nursing, 16*, 25-32. https://doi.org/10.1016/j.ecns.2017.09.010

Buttussi, F., & Chittaro, L. (2017). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics, 24*(2), 1063-1076.

Buttussi, F., & Chittaro, L. (2018, 2018-2-1). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics, 24*(2), 1063-1076. https://doi.org/10.1109/TVCG.2017.2653117

Cabaret, P.-A., Howard, T., Pacchierotti, C., Babel, M., & Marchal, M. (2022). Perception of spatialized vibrotactile impacts in a hand-held tangible for Virtual Reality. Haptics: Science, Technology, Applications: 13th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, EuroHaptics 2022, Hamburg, Germany, May 22–25, 2022, Proceedings,

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of psychophysiology*. Cambridge university press.

Calderon, D. P., Kilinc, M., Maritan, A., Banavar, J. R., & Pfaff, D. (2016). Generalized CNS arousal: An elementary force within the vertebrate nervous system. *Neuroscience and biobehavioral reviews, 68*, 167-176. https://doi.org/10.1016/j.neubiorev.2016.05.014

Cameron, L. (2017). *Step Into a Shipping Container and Emerge in Southern Italy: Virtual Reality Tourism Now Tries to Sell You a Trip to Apulia*. IEEE Computer Society. Retrieved April 2 from

Cao, L., Lin, J., & Li, N. (2019, 01/2019). A virtual reality based study of indoor fire evacuation after active or passive spatial exploration. *Computers in Human Behavior, 90*, 37-45. https://doi.org/10.1016/j.chb.2018.08.041

Carlson, P., Peters, A., Gilbert, S. B., Vance, J. M., & Luse, A. (2015, 2015-6-1). Virtual Training: Learning Transfer of Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics, 21*(6), 770-782. https://doi.org/10.1109/TVCG.2015.2393871

Cebeci, B., Celikcan, U., & Capin, T. K. (2019). A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments. *Computer Animation and Virtual Worlds, 30*(3-4), e1893. https://doi.org/https://doi.org/10.1002/cav.1893

Champseix , R. (2022). *hrv-analysis*.

Chauvel, G., Wulf, G., & Maquestiaux, F. (2015). Visual illusions can facilitate sport skill learning. *Psychonomic bulletin & review, 22*, 717-721.

Checa, D., & Bustillo, A. (2019). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 1-27.

Checa, D., & Bustillo, A. (2020, 3/2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications, 79*(9-10), 5501-5527. https://doi.org/10.1007/s11042-019-08348-9

Chinello, F., Pacchierotti, C., Bimbo, J., Tsagarakis, N. G., & Prattichizzo, D. (2017). Design and evaluation of a wearable skin stretch device for haptic guidance. *IEEE Robotics and Automation Letters, 3*(1), 524-531.

Chiossi, F., Welsch, R., Villa, S., Chuang, L., & Mayer, S. (2022). Virtual Reality Adaptation Using Electrodermal Activity to Support the User Experience. *Big Data and Cognitive Computing, 6*(2), 55.

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529-538.

Chiviacowsky, S., & Wulf, G. (2005, 2005/03/01). Self-Controlled Feedback Is Effective if It Is Based on the Learner's Performance. *Research Quarterly for Exercise and Sport, 76*(1), 42-48. https://doi.org/10.1080/02701367.2005.10599260

Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research, 66*(9), 1261-1266.

Christou, C. G., Michael-Grigoriou, D., Sokratous, D., & Tsiakoulia, M. (2018). BuzzwireVR: An Immersive Game to Supplement Fine-Motor Movement Therapy. ICAT-EGVE,

Coban, M., Bolat, Y. I., & Goksu, I. (2022, 2022/06/01/). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*, 100452. https://doi.org/https://doi.org/10.1016/j.edurev.2022.100452

Collet, C., Petit, C., Priez, A., & Dittmar, A. (2005, 2005/05/01/). Stroop color–word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological Psychology, 69*(2), 195-203. https://doi.org/https://doi.org/10.1016/j.biopsycho.2004.07.003

Collins, J., Regenbrecht, H., Langlotz, T., Said Can, Y., Ersoy, C., & Butson, R. (2019, 10/2019). Measuring Cognitive Load and Insight: A Methodology Exemplified in a Virtual Reality Learning Context. 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR),

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PloS one, 13*(2), e0191846.

Cooper, N., Millela, F., Cant, I., White, M. D., & Meyer, G. (2021). Transfer of training—Virtual reality training with augmented multisensory cues improves user experience during training and task performance in the real world. *PloS one, 16*(3), e0248225.

Cuervo, E., Chintalapudi, K., & Kotaru, M. (2018). *Creating the Perfect Illusion: What will it take to Create Life-Like Virtual Reality Headsets?* Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications, Tempe, Arizona, USA. https://doi.org/10.1145/3177102.3177115

Culbertson, H., Schorr, S. B., & Okamura, A. M. (2018). Haptics: The present and future of artificial touch sensation. *Annual Review of Control, Robotics, and Autonomous Systems, 1*, 385-409.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272-309.

Daniel Eckert, & Andrea Mower. (2020). *The Effectiveness of Virtual Reality Soft Skills Training in the Enterprise*. https://www.pwc.com/us/en/services/consulting/technology/emerging-technology/vr-study-2020.html#:~:text=In%20our%20study%2C%20VR%2Dtrained,training%20and%20have%20better%20outcomes.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2016). The Electrodermal System. In G. G. Berntson, J. T. Cacioppo, & L. G. Tassinary (Eds.), *Handbook of Psychophysiology* (4 ed., pp. 217-243). Cambridge University Press. https://doi.org/DOI: 10.1017/9781107415782.010

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron, 72*(3), 443-454.

de Lima, E. S., Silva, B. M., & Galam, G. T. (2022). Adaptive virtual reality horror games based on Machine learning and player modeling. *Entertainment Computing, 43*, 100515.

Demetriou, C., Hu, L., Smith, T. O., & Hing, C. B. (2019, 2019/12/01). Hawthorne effect on surgical studies [https://doi.org/10.1111/ans.15475]. *ANZ Journal of Surgery, 89*(12), 1567-1576. https://doi.org/https://doi.org/10.1111/ans.15475

Devigne, L., Aggravi, M., Bivaud, M., Balix, N., Teodorescu, C. S., Carlson, T., Spreters, T., Pacchierotti, C., & Babel, M. (2020). Power wheelchair navigation assistance using wearable vibrotactile haptics. *IEEE transactions on haptics, 13*(1), 52-58.

Dey, A., Chatburn, A., & Billinghurst, M. (2019, 3/2019). Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Diemer, J., Lohkamp, N., Mühlberger, A., & Zwanzger, P. (2016, 2016/01/01/). Fear and physiological arousal during a virtual height challenge—effects in patients with acrophobia and healthy controls. *Journal of anxiety disorders, 37*, 30-39. https://doi.org/https://doi.org/10.1016/j.janxdis.2015.10.007

Drey, T., Jansen, P., Fischbach, F., Frommel, J., & Rukzio, E. (2020). Towards progress assessment for adaptive hints in educational virtual reality games. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems,

Duggan, D., Kingsley, C., Mazzeo, M., & Jenkins, M. (2019). Exploring Extended Reality as a Simulation Training Tool Through Naturalistic Interactions and Enhanced Immersion. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Applications and Case Studies* (Vol. 11575, pp. 272-282). Springer International Publishing.

Durai, V. S. I., Arjunan, R., & Manivannan, M. (2019, 3/2019). The Effect of Audio and Visual Modality Based CPR Skill Training with Haptics Feedback in VR. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences, 20*(4), 155.

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., & Murray, N. (2016, 6-8 June 2016). An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist, 49*(8), 725.

Falck, A.-C., Örtengren, R., & Högberg, D. (2010). The impact of poor assembly ergonomics on product quality: A cost–benefit analysis in car manufacturing. *Human Factors and Ergonomics in Manufacturing & Service Industries, 20*(1), 24-41. https://doi.org/10.1002/hfm.20172

Falck, A. C., Örtengren, R., & Högberg, D. (2010). The impact of poor assembly ergonomics on product quality: A cost–benefit analysis in car manufacturing. *Human Factors and Ergonomics in Manufacturing & Service Industries, 20*(1), 24-41.

Feltz, D. L., Chow, G. M., & Hepler, T. J. (2008). Path analysis of self-efficacy and diving performance revisited. *Journal of Sport and Exercise Psychology, 30*(3), 401-411.

Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018, 12/2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education, 127*, 252-266. https://doi.org/10.1016/j.compedu.2018.09.002

Feuchtner, T., & Müller, J. (2017). Extending the body for interaction with reality. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,

Fisher, S. S., Wenzel, E. M., Coler, C., & McGreevy, M. W. (1988). Virtual interface environment workstations. Proceedings of the Human Factors Society Annual Meeting,

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin, 76*(5), 378.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020, 3/2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy, 34*(3), 1244-1252. https://doi.org/10.1007/s00464-019-06887-8

Freifeld, L. (2018). *Training Industry Repost*. https://trainingmag.com/

Gatti, E., Caruso, G., Bordegoni, M., & Spence, C. (2013). Can the feel of the haptic interaction modify a user's emotional state? 2013 World Haptics Conference (WHC),

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097-1114.

Gegenfurtner, A., Veermans, K., & Vauras, M. (2013). Effects of computer support, collaboration, and time lag on performance self-efficacy and transfer of training: A longitudinal meta-analysis. *Educational Research Review, 8*, 75-89.

Gerbaud, S., Gouranton, V., & Arnaldi, B. (2009). Adaptation in collaborative virtual environments for training. International conference on technologies for E-learning and digital entertainment,

Gibbs, J. K., Gillies, M., & Pan, X. (2022). A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies, 157*, 102717.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019, 2019/07/01). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology, 119*(7), 1525-1532. https://doi.org/10.1007/s00421-019-04142-5

Ginty, A. T., Oosterhoff, B. J., Young, D. A., & Williams, S. E. (2022). Effects of arousal reappraisal on the anxiety responses to stress: Breaking the cycle of negative arousal intensity and arousal interpretation. *British Journal of Psychology, 113*(1), 131-152.

GlobalData. (2022). *Virtual Reality Market Size, Share and Trends Analysis Report by End-User Type, Product Type and Region, 2021-2030*. GlobalData.

Grabowski, A., & Jankowski, J. (2015). Virtual reality-based pilot training for underground coal miners. *Safety science, 72*, 310-314.

Granato, M., Gadia, D., Maggiorini, D., & Ripamonti, L. A. (2018). Software and hardware setup for emotion recognition during video game fruition. Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good,

Grassini, S., Laumann, K., & Rasmussen Skogstad, M. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Frontiers in psychology, 11*, 1743.

Grensing-Pophal, L. (2020). How Companies Are Using AR and VR for Remote Training. Retrieved 19-05-2021, from https://hrdailyadvisor.blr.com/2020/11/18/how-companies-are-using-ar-and-vr-for-remote-training/

Grossman, R., & Salas, E. (2011). The transfer of training: what really matters. *International Journal of Training and Development, 15*(2), 103-120. https://doi.org/10.1111/j.1468-2419.2011.00373.x

Gurusamy, K. S., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2009). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane database of systematic reviews*(1).

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021, 2021/03/01). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*(1), 1-32. https://doi.org/10.1007/s40692-020-00169-2

Harris, D., Wilson, M., & Vine, S. (2019, 2019-12-21). Development and validation of a simulation workload measure: the simulation task load index (SIM-TLX). *Virtual Reality*. https://doi.org/10.1007/s10055-019-00422-9

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting,

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139-183). Elsevier.

Harvey, C., Selmanović, E., O'Connor, J., & Chahin, M. (2019, 2019-5-24). A comparison between expert and beginner learning for motor skill development in a virtual reality serious game. *The Visual Computer*. https://doi.org/10.1007/s00371-019-01702-w

Harvey, C., Selmanović, E., O'Connor, J., & Chahin, M. (2021). A comparison between expert and beginner learning for motor skill development in a virtual reality serious game. *The Visual Computer, 37*(1), 3-17.

He, Q., Cheng, X., & Cheng, Z. (2019, 11/2019). A VR-based Complex Equipment Maintenance Training System. 2019 Chinese Automation Congress (CAC),

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156-166. https://doi.org/10.1109/TITS.2005.848368

Hebert, E. P., & Coker, C. (2021). Optimizing Feedback Frequency in Motor Learning: Self-Controlled and Moderate Frequency KR Enhance Skill Acquisition. *Perceptual and Motor Skills, 128*(5), 2381-2397. https://doi.org/10.1177/00315125211036413

Heilig, M. L. (1962). Sensorama simulator. *US PAT. 3,050,870.*

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hirt, C., Holzwarth, V., Gisler, J., Schneider, J., & Kunz, A. (2019, 9/2019). Virtual Learning Environment for an Industrial Assembly Task. 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin),

Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V., Villringer, A., & Gaebler, M. (2021, 2021/10/28). Decoding subjective emotional arousal from EEG during an immersive virtual reality experience. *eLife, 10*, e64812. https://doi.org/10.7554/eLife.64812

Högberg, J., Hamari, J., & Wästlund, E. (2019, 2019/07/01). Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction, 29*(3), 619-660. https://doi.org/10.1007/s11257-019-09223-w

Holden, M. K. (2005). Virtual environments for motor rehabilitation. *CyberPsychology & Behavior, 8*(3), 187-211.

Homer, B. D., Plass, J. L., Rose, M. C., MacNamara, A. P., Pawar, S., & Ober, T. M. (2019, 2019/01/01/). Activating adolescents' "hot" executive functions in a digital game to train cognitive skills: The effects of age and prior abilities. *Cognitive Development, 49*, 20-32. https://doi.org/https://doi.org/10.1016/j.cogdev.2018.11.005

Hooper, J., Tsiridis, E., Feng, J. E., Schwarzkopf, R., Waren, D., Long, W. J., Poultsides, L., Macaulay, W., Papagiannakis, G., & Kenanidis, E. (2019). Virtual reality simulation facilitates resident training in total hip arthroplasty: a randomized controlled trial. *The Journal of arthroplasty, 34*(10), 2278-2283.

Hooper, J., Tsiridis, E., Feng, J. E., Schwarzkopf, R., Waren, D., Long, W. J., Poultsides, L., Macaulay, W., Papagiannakis, G., Kenanidis, E., Rodriguez, E. D., Slover, J., Egol, K. A., Phillips, D. P., Friedlander, S., & Collins, M. (2019, 10/2019). Virtual Reality Simulation Facilitates Resident Training in Total Hip Arthroplasty: A Randomized Controlled Trial. *The Journal of Arthroplasty, 34*(10), 2278-2283. https://doi.org/10.1016/j.arth.2019.04.002

Howard, M. C. (2017, 2017/05/01/). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior, 70*, 317-327. https://doi.org/https://doi.org/10.1016/j.chb.2017.01.013

Howard, T., Marchal, M., Lécuyer, A., & Pacchierotti, C. (2019). PUMAH: pan-tilt ultrasound mid-air haptics for larger interaction workspace in virtual reality. *IEEE transactions on haptics, 13*(1), 38-44.

Hsu, E. B., Li, Y., Bayram, J. D., Levinson, D., Yang, S., & Monahan, C. (2013). State of virtual reality based disaster preparedness and response training. *PLoS currents, 5*.

Huang, Y., Richter, E., Kleickmann, T., & Richter, D. (2022). Comparing video and virtual reality as tools for fostering interest and self-efficacy in classroom management: Results of a pre-registered experiment. *British Journal of Educational Technology*.

Huber, T., Paschold, M., Hansen, C., Wunderling, T., Lang, H., & Kneist, W. (2017, 11/2017). New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. *Surgical Endoscopy, 31*(11), 4472-4477. https://doi.org/10.1007/s00464-017-5500-6

Huber, T., Wunderling, T., Paschold, M., Lang, H., Kneist, W., & Hansen, C. (2018, 2/2018). Highly immersive virtual reality laparoscopy simulation: development and future aspects. *International Journal of Computer Assisted Radiology and Surgery, 13*(2), 281-290. https://doi.org/10.1007/s11548-017-1686-2

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering, 9*(03), 90-95.

Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS biology, 14*(9), e1002541.

Hutchinson, J. C., Sherman, T., Martinovic, N., & Tenenbaum, G. (2008). The effect of manipulated self-efficacy on perceived and sustained effort. *Journal of Applied Sport Psychology, 20*(4), 457-472.

Islam, M. S., & Lim, S. (2022). Vibrotactile feedback in virtual motor learning: A systematic review. *Applied Ergonomics, 101*, 103694.

Iván Aguilar Reyes, C., Wozniak, D., Ham, A., & Zahabi, M. (2022). An adaptive virtual reality-based training system for pilots. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Ivec, S. (2020). How Virtual Reality Can Solve COVID-19 Training Challenges: 3 Ways Immersive Learning Saves the Day. Retrieved 19-05-2021, from https://trainingindustry.com/articles/learning-technologies/how-virtual-reality-can-solve-covid-19-training-challenges-3-ways-immersive-learning-saves-the-day-spon-elearningbrothers/

Jain, S., Lee, S., Barber, S. R., Chang, E. H., & Son, Y.-J. (2020, 2020-04-02). Virtual reality based hybrid simulation for functional endoscopic sinus surgery. *IISE Transactions on Healthcare Systems Engineering, 10*(2), 127-141. https://doi.org/10.1080/24725579.2019.1692263

Janelle, C. M. (2002, 2002/01/01). Anxiety, arousal and visual attention: a mechanistic account of performance variability. *Journal of Sports Sciences, 20*(3), 237-251. https://doi.org/10.1080/026404102317284790

Jensen, L., & Konradsen, F. (2018, 7/2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*(4), 1515-1529. https://doi.org/10.1007/s10639-017-9676-0

Kassem, M., Benomran, L., & Teizer, J. (2017). Virtual environments for safety learning in construction and engineering: seeking evidence and identifying gaps for future research. *Visualization in Engineering, 5*, 1-15.

Katsis, C. D., Katertsidis, N., Ganiatsas, G., & Fotiadis, D. I. (2008). 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. https://doi.org/10.1109/TSMCA.2008.918624

Kelley, C. R. (1969). What is adaptive training? *Human Factors, 11*(6), 547-556.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993, 07/1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology, 3*(3), 203-220. https://doi.org/10.1207/s15327108ijap0303_3

Keshavarz, B., & Hecht, H. (2011, 08/2011). Validating an Efficient Method to Quantify Motion Sickness. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*(4), 415-426. https://doi.org/10.1177/0018720811403736

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002, 2002/08/09/). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters, 328*(2), 145-149. https://doi.org/https://doi.org/10.1016/S0304-3940(02)00462-7

Khan, R., Plahouras, J., Johnston, B. C., Scaffidi, M. A., Grover, S. C., & Walsh, C. M. (2019). Virtual reality simulation training in endoscopy: a Cochrane review and meta-analysis. *Endoscopy, 51*(07), 653-664.

Kim, T., & Biocca, F. (1997). Telepresence via Television: Two Dimensions of Telepresence May Have Different Connections to Memory and Persuasion.[1]. *Journal of Computer-Mediated Communication, 3*(2), 0-0. https://doi.org/https://doi.org/10.1111/j.1083-6101.1997.tb00073.x

Kleim, J. A., & Jones, T. A. (2008). Principles of Experience-Dependent Neural Plasticity: Implications for Rehabilitation After Brain Damage. *Journal of Speech, Language, and Hearing Research, 51*(1), S225-S239. https://doi.org/doi:10.1044/1092-4388(2008/018)

Klingenberg, S., Jørgensen, M. L., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology, 51*(6), 2115-2138.

Knopp, S., Lorenz, M., Pelliccia, L., & Klimant, P. (2018, 3/2018). Using Industrial Robots as Haptic Devices for VR-Training. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Kohli, L. (2010, 03/2010). Redirected touching: Warping space to remap passive haptics. 2010 IEEE Symposium on 3D User Interfaces (3DUI),

Koumaditis, K., Chinello, F., Mitkidis, P., & Karg, S. (2020a). Effectiveness of Virtual Versus Physical Training: The Case of Assembly Tasks, Trainer's Verbal Assistance, and Task Complexity. *IEEE Computer Graphics and Applications, 40*(5), 41-56.

Koumaditis, K., Chinello, F., Mitkidis, P., & Karg, S. T. (2020b, 2020). Effectiveness of Virtual vs. Physical Training: The Case of Assembly Tasks, Trainer's Verbal Assistance and Task Complexity. *IEEE Computer Graphics and Applications*, 1-1. https://doi.org/10.1109/MCG.2020.3006330

Koumaditis, K., Chinello, F., & Venckute, S. (2018). Design of a virtual reality and haptic setup linking arousals to training scenarios: A preliminary stage. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Kourtesis, P., Argelaguet, F., Vizcay, S., Marchal, M., & Pacchierotti, C. (2022). Electrotactile feedback applications for hand and arm interactions: A systematic review, meta-analysis, and future directions. *IEEE Transactions on Haptics*.

Kreimeier, J., Hammer, S., Friedmann, D., Karg, P., Bühner, C., Bankel, L., & Götzelmann, T. (2019). *Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality* Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Rhodes, Greece. https://doi.org/10.1145/3316782.3321536

Krogmeier, C., Mousas, C., & Whittinghill, D. (2019). Human–virtual character interaction: Toward understanding the influence of haptic feedback. *Computer Animation and Virtual Worlds, 30*(3-4), e1883. https://doi.org/https://doi.org/10.1002/cav.1883

Kuan, G., Morris, T., Kueh, Y. C., & Terry, P. C. (2018, 2018-February-05). Effects of Relaxing and Arousing Music during Imagery Training on Dart-Throwing Performance, Physiological Arousal Indices, and Competitive State Anxiety [Original Research]. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00014

Kuang, L., Aggravi, M., Giordano, P. R., & Pacchierotti, C. (2022). Wearable cutaneous device for applying position/location haptic feedback in navigation applications. 2022 IEEE Haptics Symposium (HAPTICS),

Lackey, S. J., Salcedo, J. N., Szalma, J. L., & Hancock, P. A. (2016, 2016-08-02). The stress and workload of virtual reality training: the effects of presence, immersion and flow. *Ergonomics, 59*(8), 1060-1072. https://doi.org/10.1080/00140139.2015.1122234

Lagos, O. (2019). *Knuckles Oculus Quest and Rift S grip*. Thingiverse. https://www.thingiverse.com/thing:3652161

Lang, Y., Wei, L., Xu, F., Zhao, Y., & Yu, L.-F. (2018, 3/2018). Synthesizing Personalized Training Programs for Improving Driving Habits via Virtual Reality. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Larmuseau, C., Cornelis, J., Lancieri, L., Desmet, P., & Depaepe, F. (2020, 2020/09/01). Multimodal learning analytics to investigate cognitive load during online problem solving [https://doi.org/10.1111/bjet.12958]. *British Journal of Educational Technology, 51*(5), 1548-1562. https://doi.org/https://doi.org/10.1111/bjet.12958

Lehikko, A. (2021). Measuring Self-Efficacy in Immersive Virtual Learning Environments: A Systematic Literature Review. *Journal of Interactive Learning Research, 32*(2), 125-146.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of psychiatric research, 45*(5), 626-629.

Levac, D. E., Huber, M. E., & Sternad, D. (2019, 2019/10/18). Learning and transfer of complex motor skills in virtual reality: a perspective review. *Journal of NeuroEngineering and Rehabilitation, 16*(1), 121. https://doi.org/10.1186/s12984-019-0587-8

Li, M., Ganni, S., Ponten, J., Albayrak, A., Rutkowski, A.-F., & Jakimowicz, J. (2020, 3/2020). Analysing usability and presence of a virtual reality operating room (VOR) simulator during laparoscopic surgery training. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Li, M., Ridzuan, M. B., Sareh, S., Seneviratne, L. D., Dasgupta, P., & Althoefer, K. (2014). Pseudo-haptics for rigid tool/soft surface interaction feedback in virtual environments. *Mechatronics, 24*(8), 1092-1100. https://doi.org/https://doi.org/10.1016/j.mechatronics.2014.07.004

Liang, Z., Zhou, K., & Gao, K. (2019, 2019). Development of Virtual Reality Serious Game for Underground Rock-Related Hazards Safety Training. *IEEE Access, 7*, 118639-118649. https://doi.org/10.1109/ACCESS.2019.2934990

Liebold, B., Brill, M., Pietschmann, D., Schwab, F., & Ohler, P. (2017, 2017/07/03). Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology, 20*(3), 477-501. https://doi.org/10.1080/15213269.2016.1206829

Liu, Z. M., Fan, X., Liu, Y., & Ye, X. d. (2022). Effects of immersive virtual reality cardiopulmonary resuscitation training on prospective kindergarten teachers' learning achievements, attitudes and self-efficacy. *British Journal of Educational Technology*.

Longo, F., Nicoletti, L., & Padovano, A. (2019, 02/2019). Emergency preparedness in industrial plants: A forward-looking solution based on industry 4.0 enabling technologies. *Computers in Industry, 105*, 99-122. https://doi.org/10.1016/j.compind.2018.12.003

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008, 06/2008). What is embodiment? A psychometric approach. *Cognition, 107*(3), 978-998. https://doi.org/10.1016/j.cognition.2007.12.004

Luong, T., Argelaguet, F., Martin, N., & Lecuyer, A. (2020, 3/2020). Introducing Mental Workload Assessment for the Design of Virtual Reality Training Scenarios. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Luvizutto, G., Bruno, A., Oliveira, S., Silva, M., & Souza, L. (2022). Development and application of an electrical buzz wire to evaluate eye-hand coordination and object control skill in children: a feasibility study. *Human Movement, 23*(2), 138-144.

Ma, M., & Bechkoum, K. (2008). Serious games for movement therapy after stroke. 2008 IEEE international conference on systems, man and cybernetics,

Mackay, C., Cox, T., Burrows, G., & Lazzerini, T. (1978). An inventory for the measurement of self-reported stress and arousal. *British Journal of Social and Clinical Psychology, 17*(3), 283-284. https://doi.org/https://doi.org/10.1111/j.2044-8260.1978.tb00280.x

Magill, R., & Anderson, D. (2016). *Motor learning and control*. McGraw-Hill Publishing New York.

Magill, R. A., & Anderson, D. I. (2018a). Memory Components, Forgetting, and Strategies. In *Motor Learning and Control: Concepts and Applications, 11e*. McGraw-Hill Education. accessphysiotherapy.mhmedical.com/content.aspx?aid=1150116546

Magill, R. A., & Anderson, D. I. (2018b). Transfer of Learning. In *Motor Learning and Control: Concepts and Applications, 11e*. McGraw-Hill Education. accessphysiotherapy.mhmedical.com/content.aspx?aid=1150116894

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021, 2021/08/01). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods, 53*(4), 1689-1696. https://doi.org/10.3758/s13428-020-01516-y

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691-707.

Makransky, G., Lilleholt, L., & Aaby, A. (2017, 2017/07/01/). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior, 72*, 276-285. https://doi.org/https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, 1-22.

Mania, K., & Chalmers, A. (2001). The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior, 4*(2), 247-264.

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics, 18*(1), 50-60, 11. https://doi.org/10.1214/aoms/1177730491

Mariani, A., Pellegrini, E., & De Momi, E. (2020). Skill-oriented and performance-driven adaptive curricula for training in robot-assisted surgery using simulators: a feasibility study. *IEEE transactions on biomedical engineering, 68*(2), 685-694.

Mariani, A., Pellegrini, E., & Momi, E. D. (2021). Skill-Oriented and Performance-Driven Adaptive Curricula for Training in Robot-Assisted Surgery Using Simulators: A Feasibility Study. *IEEE Transactions on Biomedical Engineering, 68*(2), 685-694. https://doi.org/10.1109/TBME.2020.3011867

Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Scilingo, E. P., Alcañiz, M., & Valenza, G. (2018, 2018/09/12). Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports, 8*(1), 13657. https://doi.org/10.1038/s41598-018-32063-4

MarketsAndMarkets. (2019). *Virtual Reality Market by Offering (Hardware and Software), Technology, Device Type (Head-Mounted Display, Gesture-Tracking Device), Application (Consumer, Commercial, Enterprise, Healthcare, Aerospace & Defense) and Geography - Global Forecast to 2024*. https://www.marketresearch.com/MarketsandMarkets-v3719/Virtual-Reality-Offering-Hardware-Software-12151294/

Matsas, E., & Vosniakos, G.-C. (2017, 5/2017). Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM), 11*(2), 139-153. https://doi.org/10.1007/s12008-015-0259-2

Matthews, G., & Margetts, I. (1991, 1991/06/01). Self-Report Arousal and Divided Attention: A Study of Performance Operating Characteristics. *Human Performance, 4*(2), 107-125. https://doi.org/10.1207/s15327043hup0402_2

Mayer, R. E., & Moreno, R. (1998). A cognitive theory of multimedia learning: Implications for design principles. *Journal of educational psychology, 91*(2), 358-368.

Mekbib, D. B., Han, J., Zhang, L., Fang, S., Jiang, H., Zhu, J., Roe, A. W., & Xu, D. (2020). Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. *Brain injury, 34*(4), 456-465.

Mendes, F. A. D. S., Pompeu, J. E., Lobo, A. M., da Silva, K. G., Oliveira, T. D. P., Zomignani, A. P., & Piemonte, M. E. P. (2012). Motor learning, retention and transfer after virtual-reality-based training in Parkinson's disease - effect of motor and cognitive demands of games: A longitudinal, controlled clinical study [Article]. *Physiotherapy (United Kingdom), 98*(3), 217-223. https://doi.org/10.1016/j.physio.2012.06.001

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014, 2014/01/01/). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29-40. https://doi.org/https://doi.org/10.1016/j.compedu.2013.07.033

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019, 10/2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

Mikropoulos, T. A., & Natsis, A. (2011, 2011/04/01/). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769-780. https://doi.org/https://doi.org/10.1016/j.compedu.2010.10.020

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems, 77*(12), 1321-1329.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995, 1995-12-21). Augmented reality: a class of displays on the reality-virtuality continuum. Photonics for Industrial Applications,

Mirauda, D., Capece, N., & Erra, U. (2020, 2020-01-21). Sustainable Water Management: Virtual Reality Training for Open-Channel Flow Monitoring. *Sustainability, 12*(3), 757. https://doi.org/10.3390/su12030757

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009, 2009-7-21). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moore, A. G., Hu, X., Eubanks, J. C., Aiyaz, A. A., & McMahan, R. P. (2020). A formative evaluation methodology for vr training simulations. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Moritz, S. E., Feltz, D. L., Fahrbach, K. R., & Mack, D. E. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research quarterly for exercise and sport, 71*(3), 280-294.

Morrison, T., Maciejewski, B., Giffi, C., DeRocco, E. S., McNelly, J., & Carrick, G. (2011, 2011). Boiling point? The skills gap in US manufacturing. *Deloitte and the Manufacturing Institute*.

Moses, Z. B., Luecken, L. J., & Eason, J. C. (2007, 22-26 Aug. 2007). Measuring Task-related Changes in Heart Rate Variability. 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,

Mousavi, S. M., & Iwatsuki, T. (2021). Easy task and choice: Motivational interventions facilitate motor skill learning in children. *Journal of Motor Learning and Development, 10*(1), 61-75.

Movahedi, A., Sheikh, M., Bagherzadeh, F., Hemayattalab, R., & Ashayeri, H. (2007, 2007/11/01). A Practice-Specificity-Based Model of Arousal for Achieving Peak Performance. *Journal of Motor Behavior, 39*(6), 457-462. https://doi.org/10.3200/JMBR.39.6.457-462

Mullen, G., & Davidenko, N. (2021, 03 May. 2021). Time Compression in Virtual Reality. *Timing & Time Perception, 9*(4), 377-392. https://doi.org/https://doi.org/10.1163/22134468-bja10034

Muñoz, J. E., Pope, A. T., & Velez, L. E. (2016). Integrating biocybernetic adaptation in virtual reality training concentration and calmness in target shooting. In *Physiological computing systems* (pp. 218-237). Springer.

Muñoz, J. E., Pope, A. T., & Velez, L. E. (2019). Integrating Biocybernetic Adaptation in Virtual Reality Training Concentration and Calmness in Target Shooting. In A. Holzinger, A. Pope, & H. Plácido da Silva (Eds.), *Physiological Computing Systems* (Vol. 10057, pp. 218-237). Springer International Publishing.

Murcia-Lopez, M., & Steed, A. (2018, 4/2018). A Comparison of Virtual and Physical Training Transfer of Bimanual Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics, 24*(4), 1574-1583. https://doi.org/10.1109/TVCG.2018.2793638

Nahavandi, S., Wei, L., Mullins, J., Fielding, M., Deshpande, S., Watson, M., Korany, S., Nahavandi, D., Hettiarachchi, I., Najdovski, Z., Jones, R., Mullins, A., & Carter, A. (2019). Haptically-Enabled VR-Based Immersive Fire Fighting Training Simulator. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *Intelligent Computing* (Vol. 997, pp. 11-21). Springer International Publishing.

Narciso, D., Melo, M., Raposo, J. V., Cunha, J., & Bessa, M. (2020). Virtual reality in training: an experimental study with firefighters. *Multimedia tools and applications, 79*, 6227-6245.

Naumann, A., & Hurtienne, J. (2010). Benchmarks for intuitive interaction with mobile devices. Proceedings of the 12th international conference on Human computer interaction with mobile devices and services,

Nichols, S., Haldane, C., & Wilson, J. R. (2000, 03/2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies, 52*(3), 471-491. https://doi.org/10.1006/ijhc.1999.0343

Nilsson, N. C., Serafin, S., Steinicke, F., & Nordahl, R. (2018). Natural walking in virtual reality: A review. *Computers in Entertainment (CIE), 16*(2), 1-22.

Nykänen, M., Puro, V., Tiikkaja, M., Kannisto, H., Lantto, E., Simpura, F., Uusitalo, J., Lukander, K., Räsänen, T., Heikkilä, T., & Teperi, A.-M. (2020, 2020/12/01/). Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. *Journal of Safety Research, 75*, 205-221. https://doi.org/https://doi.org/10.1016/j.jsr.2020.09.015

Okuma, T., & Kurata, T. (2016). Service field simulator: Virtual environment display system for analyzing human behavior in service fields. Serviceology for Designing the Future: Selected and Edited Papers of the 2nd International Conference on Serviceology,

Ong, N. T., Lohse, K. R., & Hodges, N. J. (2015). Manipulating target size influences perceptions of success when learning a dart-throwing skill but does not impact retention. *Frontiers in psychology*, 1378.

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., Savinainen, M., & Nygård, C.-H. (2008). Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics, 14*(3), 275-283.

Overtoom, E. M., Horeman, T., Jansen, F.-W., Dankelman, J., & Schreuder, H. W. (2019). Haptic feedback, force feedback, and force-sensing in simulation training for laparoscopy: A systematic overview. *Journal of surgical education, 76*(1), 242-261.

Owens, M. E., & Beidel, D. C. (2015, 2015/06/01). Can Virtual Reality Effectively Elicit Distress Associated with Social Anxiety Disorder? *Journal of Psychopathology and Behavioral Assessment, 37*(2), 296-305. https://doi.org/10.1007/s10862-014-9454-x

Pacchierotti, C. (2022). *Cutaneous haptic feedback for robotics and Virtual Reality* Université de Rennes 1 (UR1), Rennes, FRA.].

Pacchierotti, C., Sinclair, S., Solazzi, M., Frisoli, A., Hayward, V., & Prattichizzo, D. (2017, Oct-Dec). Wearable Haptic Systems for the Fingertip and the Hand: Taxonomy, Review, and Perspectives. *IEEE Trans Haptics, 10*(4), 580-600. https://doi.org/10.1109/TOH.2017.2689006

Pakarinen, T., Pietilä, J., & Nieminen, H. (2019, 23-27 July 2019). Prediction of Self-Perceived Stress and Arousal Based on Electrodermal Activity[*]. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

Panariello, D., Caporaso, T., Grazioso, S., Di Gironimo, G., Lanzotti, A., Knopp, S., Pelliccia, L., Lorenz, M., & Klimant, P. (2019, 2/2019). Using the KUKA LBR iiwa Robot as Haptic Device for Virtual Reality Training of Hip Replacement Surgery. 2019 Third IEEE International Conference on Robotic Computing (IRC),

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226-241. https://doi.org/https://doi.org/10.1111/jcal.12482

Pascua, L. A., Wulf, G., & Lewthwaite, R. (2015). Additive benefits of external focus and enhanced performance expectancy for motor learning. *Journal of sports sciences, 33*(1), 58-66.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci., 28*, 377-401.

Patle, D. S., Manca, D., Nazir, S., & Sharma, S. (2019, 9/2019). Operator training simulators in virtual reality environment for process operators: a review. *Virtual Reality, 23*(3), 293-311. https://doi.org/10.1007/s10055-018-0354-3

Pavlidis, I., Zavlin, D., Khatri, A. R., Wesley, A., Panagopoulos, G., & Echo, A. (2019, 2019/02/11). Absence of Stressful Conditions Accelerates Dexterous Skill Acquisition in Surgery. *Scientific reports, 9*(1), 1747. https://doi.org/10.1038/s41598-019-38727-z

Pellas, N., Fotaris, P., Kazanidis, I., & Wells, D. (2019, 2019/12/01). Augmenting the learning experience in primary and secondary school education: a systematic review of recent trends in augmented reality game-based learning. *Virtual Reality, 23*(4), 329-346. https://doi.org/10.1007/s10055-018-0347-2

Pelliccia, L., Lorenz, M., Heyde, C.-E., Kaluschke, M., Klimant, P., Knopp, S., Schleifenbaum, S., Rotsch, C., Weller, R., & Werner, M. (2020). A cadaver-based biomechanical model of acetabulum reaming for surgical virtual reality training simulators. *Scientific reports, 10*(1), 1-12.

Pérez, L., Diez, E., Usamentiaga, R., & García, D. F. (2019, 08/2019). Industrial robot control and operator training using virtual reality interfaces. *Computers in Industry, 109*, 114-120. https://doi.org/10.1016/j.compind.2019.05.001

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning, 34*(4), 397-408. https://doi.org/https://doi.org/10.1111/jcal.12271

Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review, 31*, 339-359.

Polce, E. M., Kunze, K. N., Williams, B. T., Krivicich, L. M., Maheshwer, B., Beletsky, A., Cohn, M. R., Kogan, M., & Chahla, J. (2020). Efficacy and validity of orthopaedic simulators in surgical training: a systematic review and meta-analysis of randomized controlled trials. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 28*(24), 1027-1040.

Pollard, K. A., Oiknine, A. H., Files, B. T., Sinatra, A. M., Patton, D., Ericson, M., Thomas, J., & Khooshabeh, P. (2020, 2020-2-28). Level of immersion affects spatial learning in virtual environments: results of a three-condition within-subjects study with long intersession intervals. *Virtual Reality*. https://doi.org/10.1007/s10055-019-00411-y

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology, 17*(3), 715-734.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

Prabhu, A., Smith, W., Yurko, Y., Acker, C., & Stefanidis, D. (2010, 2010/05/01/). Increased stress levels may explain the incomplete transfer of simulator-acquired skill to the operating room. *Surgery, 147*(5), 640-645. https://doi.org/https://doi.org/10.1016/j.surg.2010.01.007

Pucher, P. H., Batrick, N., Taylor, D., Chaudery, M., Cohen, D., & Darzi, A. (2014). Virtual-world hospital simulation for real-world disaster response: design and validation of a virtual reality simulator for mass casualty incident management. *Journal of trauma and acute care surgery, 77*(2), 315-321.

Pulijala, Y., Ma, M., Pears, M., Peebles, D., & Ayoub, A. (2018, 05/2018). Effectiveness of Immersive Virtual Reality in Surgical Training—A Randomized Control Trial. *Journal of Oral and Maxillofacial Surgery, 76*(5), 1065-1072. https://doi.org/10.1016/j.joms.2017.10.002

Putze, S., Alexandrovsky, D., Putze, F., Höffner, S., Smeddinck, J. D., & Malaka, R. (2020). *Breaking The Experience: Effects of Questionnaires in VR User Studies* Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376144

Quick, J. A., Bukoski, A. D., Doty, J., Bennett, B. J., Crane, M., & Barnes, S. L. (2017). Objective measurement of clinical competency in surgical education using electrodermal activity. *Journal of Surgical Education, 74*(4), 674-680.

Radhakrishnan, U., Blindu, A., Chinello, F., & Koumaditis, K. (2021). Investigating motor skill training and user arousal levels in VR: Pilot Study and Observations. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW),

Radhakrishnan, U., Chinello, F., & Koumaditis, K. (2021, 27 March-1 April 2021). Immersive Virtual Reality Training : Three Cases from the Danish Industry. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW),

Radhakrishnan, U., Chinello, F., & Koumaditis, K. (2022, 2022/11/14). Investigating the effectiveness of immersive VR skill training and its link to physiological arousal. *Virtual Reality*. https://doi.org/10.1007/s10055-022-00699-3

Radhakrishnan, U., Koumaditis, K., & Chinello, F. (2021). A systematic review of immersive virtual reality for industrial skills training. *Behaviour & Information Technology, 40*(12), 1310-1339.

Radhakrishnan, U., Koumaditis, K., & Chinello, F. (2022). Investigating Haptic Feedback and Arousal for Motor Skill Training in Virtual Reality. Haptics: Science, Technology, Applications: 13th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, EuroHaptics 2022, Hamburg, Germany, May 22–25, 2022, Proceedings,

Radhakrishnan, U., Kuang, L., Koumaditis, K., Chinello, F., & Pacchierotti, C. (2023). *Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task*.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778.

Radkowski, R., & Raul, S. (2019). Impact of Foveated Rendering on Procedural Task Training. In J. Y. C. Chen & G. Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality. Multimodal Interaction* (Vol. 11574, pp. 258-267). Springer International Publishing.

Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015, 2015-7-1). Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics, 21*(7), 794-807. https://doi.org/10.1109/TVCG.2015.2403312

Ragan, E. D., Scerbo, S., Bacim, F., & Bowman, D. A. (2017, 2017-8-1). Amplified Head Rotation in Virtual Reality and the Effects on 3D Search, Training Transfer, and Spatial Orientation. *IEEE Transactions on Visualization and Computer Graphics, 23*(8), 1880-1895. https://doi.org/10.1109/TVCG.2016.2601607

Rahman, Y., Asish, S. M., Fisher, N. P., Bruce, E. C., Kulshreshth, A. K., & Borst, C. W. (2020, 3/2020). Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Ramírez-Fernández, C., Morán, A. L., & García-Canseco, E. (2015). Haptic feedback in motor hand virtual therapy increases precision and generates less mental workload. 2015 9th international conference on pervasive computing technologies for healthcare (PervasiveHealth),

Rangarajan, K., Davis, H., & Pucher, P. H. (2020, 2020/03/01/). Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool? *Journal of Surgical Education, 77*(2), 337-347. https://doi.org/https://doi.org/10.1016/j.jsurg.2019.09.006

Read, J. C., Begum, S. F., McDonald, A., & Trowbridge, J. (2013). The binocular advantage in visuomotor tasks involving tools. *i-Perception, 4*(2), 101-110.

Reason, J., & Hobbs, A. (2017). *Managing maintenance error: a practical guide*. CRC Press.

Renganayagalu, S. k., Mallam, S. C., & Nazir, S. (2021, 2021/12/01). Effectiveness of VR Head Mounted Displays in Professional Training: A Systematic Review. *Technology, Knowledge and Learning, 26*(4), 999-1041. https://doi.org/10.1007/s10758-020-09489-9

Richard, G., Pietrzak, T., Argelaguet, F., Lécuyer, A., & Casiez, G. (2022). Within or between? comparing experimental designs for virtual embodiment studies. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior, 10*(1), 45-56.

Roldán, J. J., Crespo, E., Martín-Barrio, A., Peña-Tapia, E., & Barrientos, A. (2019, 10/2019). A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining. *Robotics and Computer-Integrated Manufacturing, 59*, 305-316. https://doi.org/10.1016/j.rcim.2019.05.004

Rose, T., Nam, C. S., & Chen, K. B. (2018). Immersion of virtual reality for rehabilitation - Review [Review]. *Applied Ergonomics, 69*, 153-161. https://doi.org/10.1016/j.apergo.2018.01.009

Rosenqvist, O., & Skans, O. N. (2015). Confidence enhanced performance?–The causal effects of success on future performance in professional golf tournaments. *Journal of Economic Behavior & Organization, 117*, 281-295.

Rubin, D. C., & Talarico, J. M. (2009). A comparison of dimensional models of emotion: evidence from emotions, prototypical events, autobiographical memories, and words. *Memory (Hove, England), 17*(8), 802-808. https://doi.org/10.1080/09658210903130764

Ruxton, G. D., & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution, 1*(2), 114-117.

[Record #8 is using a reference type undefined in this output style.]

Saemi, E., Porter, J. M., Ghotbi-Varzaneh, A., Zarghami, M., & Maleki, F. (2012). Knowledge of results after relatively good trials enhances self-efficacy and motor learning. *Psychology of Sport and Exercise, 13*(4), 378-382.

Sainsbury, B., Łącki, M., Shahait, M., Goldenberg, M., Baghdadi, A., Cavuoto, L., Ren, J., Green, M., Lee, J., Averch, T. D., & Rossa, C. (2020, 2020-1-14). Evaluation of a Virtual Reality Percutaneous Nephrolithotomy (PCNL) Surgical Simulator. *Frontiers in Robotics and AI, 6*, 145. https://doi.org/10.3389/frobt.2019.00145

Sakowitz, S. M., Inglehart, M. R., Ramaswamy, V., Edwards, S., Shoukri, B., Sachs, S., & Kim-Berman, H. (2019, 2019-12-4). A comparison of two-dimensional prediction tracing and a virtual reality patient methods for diagnosis and treatment planning of orthognathic cases in dental students: a randomized preliminary study. *Virtual Reality*. https://doi.org/10.1007/s10055-019-00413-w

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological science in the public interest, 13*(2), 74-101.

Sampath, H. A., Indurkhya, B., Lee, E., & Bae, Y. (2015). *Towards Multimodal Affective Feedback: Interaction between Visual and Haptic Modalities* Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea. https://doi.org/10.1145/2702123.2702288

Sattelmayer, M., Elsig, S., Hilfiker, R., & Baer, G. (2016, 2016/01/15). A systematic review and meta-analysis of selected motor learning principles in physiotherapy and medical education. *BMC Medical Education, 16*(1), 15. https://doi.org/10.1186/s12909-016-0538-z

Schachinger, H., Blumenthal, T. D., Richter, S., Savaskan, E., Wirz-Justice, A., & Kräuchi, K. (2008, 2008/07/01/). Melatonin reduces arousal and startle responsiveness without influencing startle habituation or affective startle modulation in young women. *Hormones and Behavior, 54*(2), 258-262. https://doi.org/https://doi.org/10.1016/j.yhbeh.2008.03.013

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001, 06/2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments, 10*(3), 266-281. https://doi.org/10.1162/105474601300343603

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In *Handbook of motivation at school* (pp. 34-54). Routledge.

Schwarz, S., Regal, G., Kempf, M., & Schatz, R. (2020). Learning Success in Immersive Virtual Reality Training Environments: Practical Evidence from Automotive Assembly. Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior, 29*(3), 1150-1158.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery, 236*(4), 458.

Shaffer, F., & Ginsberg, J. P. (2017, 2017-September-28). An Overview of Heart Rate Variability Metrics and Norms [Review]. *Frontiers in Public Health, 5*. https://doi.org/10.3389/fpubh.2017.00258

Shafti, A., Lazpita, B. U., Elhage, O., Wurdemann, H. A., & Althoefer, K. (2016). Analysis of comfort and ergonomics for clinical work environments. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

Shakur, S. F., Luciano, C. J., Kania, P., Roitberg, B. Z., Banerjee, P. P., Slavin, K. V., Sorenson, J., Charbel, F. T., & Alaraj, A. (2015, 2015-09-01). Usefulness of a Virtual Reality Percutaneous Trigeminal Rhizotomy Simulator in Neurosurgical Training. *Operative Neurosurgery, 11*(3), 420-425. https://doi.org/10.1227/NEU.0000000000000853

Shamsuzzoha, A., Toshev, R., Vu Tuan, V., Kankaanpaa, T., & Helo, P. (2019, 2019-06-10). Digital factory – virtual reality environments for industrial training and maintenance. *Interactive Learning Environments*, 1-24. https://doi.org/10.1080/10494820.2019.1628072

Sherman, W. R., & Craig, A. B. (2003). Understanding virtual reality. *San Francisco, CA: Morgan Kauffman*.

Shi, Y., Du, J., & Worthy, D. A. (2020). The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment. *Automation in Construction, 119*, 103367.

Shi, Y., Zhu, Y., Mehta, R. K., & Du, J. (2020). A neurophysiological approach to assess training outcome under stress: A virtual reality experiment of industrial shutdown maintenance using Functional Near-Infrared Spectroscopy (fNIRS). *Advanced Engineering Informatics, 46*, 101153.

Shu, Y., Huang, Y.-Z., Chang, S.-H., & Chen, M.-Y. (2019, 2019/12/01). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality, 23*(4), 437-446. https://doi.org/10.1007/s10055-018-0376-x

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review, 20*, 21-53.

Simeone, A. L., Speicher, M., Molnar, A., Wilde, A., & Daiber, F. (2019, 2019-10-19). LIVE: the Human Role in Learning in Immersive Virtual Environments. SUI '19: Symposium on Spatial User Interaction,

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: what we know and where we need to go. *Psychological bulletin, 137*(3), 421.

Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality, 2*, 647997.

Škola, F., Tinková, S., & Liarokapis, F. (2019, 2019-9-26). Progressive Training for Motor Imagery Brain-Computer Interfaces Using Gamification and Virtual Reality Embodiment. *Frontiers in Human Neuroscience, 13*, 329. https://doi.org/10.3389/fnhum.2019.00329

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology, 109*(3), 431-433.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., & Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence, 15*(5), 553-569.

Song, H., Kim, T., Kim, J., Ahn, D., & Kang, Y. (2021). Effectiveness of VR crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction, 122*, 103506.

Sportillo, D., Avveduto, G., Tecchia, F., & Carrozzino, M. (2015). Training in VR: a preliminary study on learning assembly/disassembly sequences. Augmented and Virtual Reality: Second International Conference, AVR 2015, Lecce, Italy, August 31-September 3, 2015, Proceedings 2,

Srinivasan, M. A., & Basdogan, C. (1997). Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics, 21*(4), 393-404. https://doi.org/10.1016/s0097-8493(97)00030-7

Srivastava, P., Rimzhim, A., Vijay, P., Singh, S., & Chandra, S. (2019, 2019-7-9). Desktop VR Is Better Than Non-ambulatory HMD VR for Spatial Learning. *Frontiers in Robotics and AI, 6*, 50. https://doi.org/10.3389/frobt.2019.00050

Stake, R. E. (1995). *The art of case study research*. sage.

Sternad, D. (2018, 2018/04/01/). It's not (only) the mean that matters: variability, noise and exploration in skill learning. *Current Opinion in Behavioral Sciences, 20*, 183-195. https://doi.org/https://doi.org/10.1016/j.cobeha.2018.01.004

Stevens, D., Anderson, D. I., O'Dwyer, N. J., & Williams, A. M. (2012). Does self-efficacy mediate transfer effects in the learning of easy and difficult motor skills? *Consciousness and cognition, 21*(3), 1122-1128.

Storbeck, J., & Clore, G. L. (2008). Affective Arousal as Information: How Affective Arousal Influences Judgments, Learning, and Memory. *Social and Personality Psychology Compass, 2*(5), 1824-1843. https://doi.org/https://doi.org/10.1111/j.1751-9004.2008.00138.x

Suh, A., & Prophet, J. (2018, 09/2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior, 86*, 77-90. https://doi.org/10.1016/j.chb.2018.04.019

Sun, R., Wu, Y. J., & Cai, Q. (2019, 12/2019). The effect of a virtual reality learning environment on learners' spatial ability. *Virtual Reality, 23*(4), 385-398. https://doi.org/10.1007/s10055-018-0355-2

Sutarto, A. P., Wahab, M. N. A., & Zin, N. M. (2010). Heart Rate Variability (HRV) biofeedback: A new training approach for operator's performance enhancement. *Journal of industrial engineering and management, 3*(1), 176-198.

Sutherland, I. E. (1968). A head-mounted three dimensional display. Proceedings of the December 9-11, 1968, fall joint computer conference, part I, Syrjämäki, A. H., Isokoski, P., Surakka, V., Pasanen, T. P., & Hietanen, J. K. (2020, 2020/11/01/). Eye contact in virtual reality – A psychophysiological study. *Computers in Human Behavior, 112*, 106454. https://doi.org/https://doi.org/10.1016/j.chb.2020.106454

Tabbron, G., & Yang, J. (1997, 7/1997). The interaction between technical and vocational education and training (TVET) and economic development in advanced countries. *International Journal of Educational Development, 17*(3), 323-334. https://doi.org/10.1016/S0738-0593(96)00072-7

Tai, K.-H., Hong, J.-C., Tsai, C.-R., Lin, C.-Z., & Hung, Y.-H. (2022, 2022/06/01/). Virtual reality for car-detailing skill development: Learning outcomes of procedural accuracy and performance quality predicted by VR self-efficacy, VR using anxiety, VR learning interest and flow experience. *Computers & Education, 182*, 104458. https://doi.org/https://doi.org/10.1016/j.compedu.2022.104458

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics, 39*(5), 740-748.

Terkildsen, T., & Makransky, G. (2019, 2019/06/01/). Measuring presence in video games: An investigation of the potential use of physiological measures as indicators of presence. *International Journal of Human-Computer Studies, 126*, 64-80. https://doi.org/https://doi.org/10.1016/j.ijhcs.2019.02.006

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. *BMC medical research methodology, 10*, 1-10.

Tian, F., Hua, M., Zhang, W., Li, Y., & Yang, X. (2021). Emotional arousal in 2D versus 3D virtual reality environments. *PLOS ONE, 16*(9), e0256211. https://doi.org/10.1371/journal.pone.0256211

Torres-Guerrero, F., Neira-Tovar, L., & Torres-Treviño, L. (2019). An Introductive Training for Welding Workshop: A Biometric Evaluation Using Virtual Reality Scenes to Improve Practice. In J. Trojanowska, O. Ciszak, J. M. Machado, & I. Pavlenko (Eds.), *Advances in Manufacturing II* (pp. 319-331). Springer International Publishing.

Turner, M., & Jones, M. (2018). Arousal control in sport. In *Oxford Research Encyclopedia of Psychology*.

Ünal, A. B., de Waard, D., Epstude, K., & Steg, L. (2013, 2013/11/01/). Driving with music: Effects on arousal and performance. *Transportation Research Part F: Traffic Psychology and Behaviour, 21*, 52-65. https://doi.org/https://doi.org/10.1016/j.trf.2013.09.004

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoper. Virtual Environ., 9*(5), 497–503. https://doi.org/10.1162/105474600566989

Vallat, R. (2018). Pingouin: statistics in Python. *J. Open Source Softw., 3*(31), 1026.

van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior, 106*(2), 298-304.

Van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education, 44*(1), 85-93. https://doi.org/https://doi.org/10.1111/j.1365-2923.2009.03498.x

Vaughan, N., Gabrys, B., & Dubey, V. N. (2016, 11/2016). An overview of self-adaptive technologies within virtual reality training. *Computer Science Review, 22*, 65-87. https://doi.org/10.1016/j.cosrev.2016.09.001

Vaughan, N., John, N., & Rees, N. (2019, 10/2019). ParaVR: Paramedic Virtual Reality Training Simulator. 2019 International Conference on Cyberworlds (CW),

Ventura, S., Cebolla, A., Latorre, J., Escrivá-Martínez, T., Llorens, R., & Baños, R. (2021, 2021/08/15). The benchmark framework and exploratory study to investigate the feasibility of 360-degree video-based virtual reality to induce a full body illusion. *Virtual Reality*. https://doi.org/10.1007/s10055-021-00567-6

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., Vázquez-Baeza, Y., & SciPy, C. (2020, 2020/03/01). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2

Vorderstrasse, A. A., Melkus, G. D., Pan, W., Lewinski, A. A., & Johnson, C. M. (2015). Diabetes learning in virtual environments: Testing the efficacy of self-management training and support in virtual environments (randomized controlled trial protocol) [Article]. *Nursing Research, 64*(6), 485-493. https://doi.org/10.1097/NNR.0000000000000128

vrchewal. (2020). *Measurements*. Unity. https://assetstore.unity.com/packages/tools/utilities/measurements-111690

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics, 24*(4), 1643-1652.

Wang, C.-A., Baird, T., Huang, J., Coutinho, J. D., Brien, D. C., & Munoz, D. P. (2018, 2018-December-03). Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task [Original Research]. *Frontiers in Neurology, 9*. https://doi.org/10.3389/fneur.2018.01029

Wang, L., Du, S., Liu, H., Yu, J., Cheng, S., & Xie, P. (2017). A virtual rehabilitation system based on EEG-EMG feedback control. 2017 Chinese Automation Congress (CAC),

Wang, P., Min, H., Bai, X., Billinghurst, M., Zhang, S., Han, D., Lv, H., He, W., Yan, Y., & Zhang, X. (2019, 10/2019). An MR Remote Collaborative Platform Based on 3D CAD Models for Training in Industry. 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct),

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software, 6*(60), 3021.

Weber, B., Sagardia, M., Hulin, T., & Preusche, C. (2013). Visual, vibrotactile, and force feedback of collisions in virtual environments: effects on performance, mental workload and spatial orientation. Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 5,

White, R. T., & Arzi, H. J. (2005). Longitudinal studies: Designs, validity, practicality, and value. *Research in science education, 35*(1), 137-149.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin, 1*(6), 80-83. https://doi.org/10.2307/3001968

Winther, F., Ravindran, L., Svendsen, K. P., & Feuchtner, T. (2020, 3/2020). Design and Evaluation of a VR Training Simulation for Pump Maintenance Based on a Use Case at Grundfos. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Witmer, B. G., & Singer, M. J. (1998, 06/1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225-240. https://doi.org/10.1162/105474698565686

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., & Parsons, T. D. (2010). Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task. *IEEE Transactions on Affective Computing, 1*(2), 109-118. https://doi.org/10.1109/T-AFFC.2010.12

Wu, S.-H., Huang, C.-C., Huang, S.-S., Yang, Y.-Y., Liu, C.-W., Shulruf, B., & Chen, C.-H. (2020, 2020-01-20). Effects of virtual reality training on decreasing the rates of needlestick or sharp injury in new-coming medical and nursing interns in Taiwan. *Journal of Educational Evaluation for Health Professions, 17*, 1. https://doi.org/10.3352/jeehp.2020.17.1

Wulf, G., Chiviacowsky, S., & Cardozo, P. L. (2014). Additive benefits of autonomy support and enhanced expectancies for motor learning. *Human movement science, 37*, 12-20.

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic bulletin & review, 23*, 1382-1414.

Wulf, G., Shea, C., & Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. *Medical education, 44*(1), 75-84.

Wulfert, E., Roland, B. D., Hartley, J., Wang, N., & Franco, C. (2005). Heart rate arousal and excitement in gambling: winners versus losers. *Psychology of Addictive Behaviors, 19*(3), 311.

Xiao, S., Ye, X., Guo, Y., Gao, B., & Long, J. (2020). Transfer of Coordination Skill to the Unpracticed Hand in Immersive Environments. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR),

Xiao, X., Zhao, S., Meng, Y., Soghier, L., Zhang, X., & Hahn, J. (2020, 3/2020). A Physics-based Virtual Reality Simulation Framework for Neonatal Endotracheal Intubation. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., & Akdere, M. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality, 2*, 49.

Xin, B., Huang, X., Wan, W., Lv, K., Hu, Y., Wang, J., Li, S., Zou, W., Xiao, J., & Liu, T. (2020, 5/2020). The efficacy of immersive virtual reality surgical simulator training for pedicle screw placement: a randomized double-blind controlled trial. *International Orthopaedics, 44*(5), 927-934. https://doi.org/10.1007/s00264-020-04488-y

Yerkes, R. M., & Dodson, J. D. (1908, 11/1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482. https://doi.org/10.1002/cne.920180503

Yildiz, E., Melo, M., Moller, C., & Bessa, M. (2019, 11/2019). Designing Collaborative and Coordinated Virtual Reality Training Integrated with Virtual and Physical Factories. 2019 International Conference on Graphics and Interaction (ICGI),

Yin, J., Arfaei, N., MacNaughton, P., Catalano, P. J., Allen, J. G., & Spengler, J. D. (2019). Effects of biophilic interventions in office on stress reaction and cognitive function: A randomized crossover study in virtual reality. *Indoor Air, 29*(6), 1028-1039. https://doi.org/https://doi.org/10.1111/ina.12593

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010, Oct). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc, 5*(5), 267-271. https://doi.org/10.1097/SIH.0b013e3181e3f329

Zahabi, M., & Abdul Razak, A. M. (2020, 2020-3-13). Adaptive virtual reality-based training: a systematic literature review and framework. *Virtual Reality*. https://doi.org/10.1007/s10055-020-00434-w

Zhao, R., Aqlan, F., Elliott, L. J., & Lum, H. C. (2019, 2019-08-26). Developing a virtual reality game for manufacturing education. FDG '19: The Fourteenth International Conference on the Foundations of Digital Games,

Zhou, M., Jones, D., Schwaitzberg, S., & Cao, C. (2007). Role of haptic feedback and cognitive load in surgical skill acquisition. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Zhou, Z., Jiang, S., Yang, Z., & Zhou, L. (2019, 12/2019). Personalized planning and training system for brachytherapy based on virtual reality. *Virtual Reality, 23*(4), 347-361. https://doi.org/10.1007/s10055-018-0350-7

Zikas, P., Protopsaltis, A., Lydatakis, N., Kentros, M., Geronikolakis, S., Kateros, S., Kamarianakis, M., Evangelou, G., Filippidis, A., & Grigoriou, E. (2023). MAGES 4.0: Accelerating the world's transition to VR training and democratizing the authoring of the medical metaverse. *IEEE Computer Graphics and Applications, 43*(2), 43-56.

Zinchenko, Y. P., Khoroshikh, P. P., Sergievich, A. A., Smirnov, A. S., Tumyalis, A. V., Kovalev, A. I., Gutnikov, S. A., & Golokhvast, K. S. (2020). Virtual reality is more efficient in learning human heart anatomy especially for subjects with low baseline knowledge. *New Ideas in Psychology, 59*. https://doi.org/10.1016/j.newideapsych.2020.100786

Zou, H., Li, N., & Cao, L. (2016, 2016-07-21). Immersive Virtual Environments for Investigating Building Emergency Evacuation Behaviors: A Feasibility Study. 33th International Symposium on Automation and Robotics in Construction,

Co-Author Statements

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Training, Quickly and Accurately: A Controlled, Preregistered Experiment on Self-Efficacy and Performance in Adaptive Virtual Training
Authors:	Lasse F. Lui, Unnikrishnan Radhakrishnan, Francesco Chinello, Konstantinos Koumaditis

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☐ In preparation ☒

If published, state full reference:

If accepted or submitted, state journal:

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	C
3. Involvement in the experimental work/clinical studies/data collection	C
4. Interpretation of the results	D
5. Writing of the first draft of the manuscript	D
6. Finalization of the manuscript and submission	D

Signatures of the co-authors

Date	Name	Signature
30-04-2023	Unnikrishnan Radhakrishnan	
30-04-2023	Konstantinos Koumaditis	KK
30-04-2023	Francesco Chinello	Francesco Chinello
30-04-2023	Lasse F. Lui	Lasse Lui Frandsen

In case of further co-authors please attach appendix

Date: 30-04-2023

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS | SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Adaptive Immersive VR Training Based on Performance and Self-Efficacy
Authors:	Lasse F. Lui, Unnikrishnan Radhakrishnan, Francesco Chinello, Konstantinos Koumaditis

The article/manuscript is: Published ☐ Accepted ☒ Submitted ☐ In preparation ☐

If published, state full reference:

If accepted or submitted, state journal: IEEE VR 2023

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	C
3. Involvement in the experimental work/clinical studies/data collection	E
4. Interpretation of the results	E
5. Writing of the first draft of the manuscript	D
6. Finalization of the manuscript and submission	B

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	[signature]
29-03-2023	Konstantinos Koumaditis	KK
29-03-2023	Francesco Chinello	Francesco Chinello
29-03-2023	Lasse F. Lui	Lasse Lui Frandsen

In case of further co-authors please attach appendix

Date: 29-03-2023

[signature]

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS | SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Haptic Feedback, Performance and Arousal:A Comparison Study in an Immersive VR Motor Skill Training Task
Authors:	Unnikrishnan Radhakrishnan, Lisheng Kuang, Francesco Chinello, Konstantinos Koumaditis, Claudio Pacchierotti

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☒ In preparation ☐

If published, state full reference:

If accepted or submitted, state journal: IEEE Transactions on Haptics

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	C
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	B
6. Finalization of the manuscript and submission	B

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	
	Konstantinos Koumaditis	KK
	Francesco Chinello	Francesco Chinello
	Lisheng Kuang	Lisheng Kuang
	Claudio Pacchierotti	Claudio Pacchierotti

In case of further co-authors please attach appendix

Date: 29-03-2023

____
Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Investigating motor skill training and user arousal levels in VR : Pilot Study and Observations
Authors:	Unnikrishnan Radhakrishnan, Alin Blindu, Francesco Chinello, Konstantinos Koumaditis

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Radhakrishnan, U., Blindu, A., Chinello, F., & Koumaditis, K. (2021, March). Investigating motor skill training and user arousal levels in VR: pilot study and observations. In 2021 IEEE conference on virtual reality and 3d user interfaces

If accepted or submitted, state journal: IEEE Conference on Virtual Reality and 3D User Interfaces

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	B
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	A
6. Finalization of the manuscript and submission	A

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	[signature]
29-03-2023	Konstantinos Koumaditis	KK
29-03-2023	Francesco Chinello	Francesco Chinello
30-03-2023	Alin Blindu	Alin Blindu

In case of further co-authors please attach appendix

Date: 29-03-2023

[signature]

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Investigating the effectiveness of immersive VR skill training and its link to physiological arousal
Authors:	Unnikrishnan Radhakrishnan, Francesco Chinello, Konstantinos Koumaditis

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Radhakrishnan, U., Chinello, F. & Koumaditis, K. Investigating the effectiveness of immersive VR skill training and its link to physiological arousal. Virtual Reality (2022). https://doi.org/10.1007/s10055-022-00699-3

If accepted or submitted, state journal: Virtual Reality

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	B
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	B
6. Finalization of the manuscript and submission	B

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	
29-03-2023	Konstantinos Koumaditis	KK
29-03-2023	Francesco Chinello	Francesco Chinello

In case of further co-authors please attach appendix

Date: 29-03-2023

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	Immersive virtual reality training: Three cases from the danish industry
Authors:	Unnikrishnan Radhakrishnan, Francesco Chinello, Konstantinos Koumaditis

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Radhakrishnan, U., Chinello, F., & Koumaditis, K. (2021, March). Immersive virtual reality training: three cases from the danish industry. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 1-5). IEEE.

If accepted or submitted, state journal: IEEE Conference on Virtual Reality and 3D User Interfaces

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	B
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	B
6. Finalization of the manuscript and submission	B

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	
29-03-2023	Konstantinos Koumaditis	KK
29-03-2023	Francesco Chinello	Francesco Chinello

In case of further co-authors please attach appendix

Date: 29-03-2023

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship*

Full name of the PhD student: Unnikrishnan Radhakrishnan

This declaration concerns the following article/manuscript:

Title:	A systematic review of immersive virtual reality for industrial skills training
Authors:	Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, Francesco Chinello

The article/manuscript is: Published ☒ Accepted ☐ Submitted ☐ In preparation ☐

If published, state full reference: Radhakrishnan, U., Koumaditis, K., & Chinello, F. (2021). A systematic review of immersive virtual reality for industrial skills training. Behaviour & Information Technology, 40(12), 1310-1339.

If accepted or submitted, state journal: Behaviour & Information Technology

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	C
2. Planning of the experiments/methodology design and development	C
3. Involvement in the experimental work/clinical studies/data collection	B
4. Interpretation of the results	C
5. Writing of the first draft of the manuscript	C
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
29-03-2023	Unnikrishnan Radhakrishnan	
29-03-2023	Konstantinos Koumaditis	KK
29-03-2023	Francesco Chinello	Francesco Chinello

In case of further co-authors please attach appendix

Date: 29-03-2023

Signature of the PhD student

*As per policy the co-author statement will be published with the dissertation.